

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Владимирский государственный университет имени Александра
Григорьевича и Николая Григорьевича Столетовых»



На правах рукописи

Бредихин Александр Юрьевич
**«РАЗРАБОТКА МЕТОДОВ И СРЕДСТВ ПОВЫШЕНИЯ
РАЗБОРЧИВОСТИ И РАСПОЗНАВАНИЯ ЗВУКОВЫХ СИГНАЛОВ
ДЛЯ ДОСТУПА СЛАБОСЛЫШАЮЩИХ АБОНЕНТОВ К РЕСУРСАМ
СЕТЕЙ, СИСТЕМ И УСТРОЙСТВ ТЕЛЕКОММУНИКАЦИЙ»**

Специальность - 2.2.15

«Системы, сети и устройства телекоммуникаций»

Диссертация на соискание ученой степени кандидата технических наук

Научный руководитель:
Самойлов Александр Георгиевич,
доктор технических наук, профессор

Владимир - 2025 г.

СОДЕРЖАНИЕ

ОГЛАВЛЕНИЕ		2-3
Введение		4-12
Глава 1	Исследование проблем разборчивости звуковых сигналов для доступа слабослышащих абонентов к ресурсам сетей, систем и устройств телекоммуникаций	13-39
1.1	Текущее состояние проблемы разборчивости звуковых сигналов для слабослышащих абонентов	13-20
1.2	Анализ способов повышения разборчивости речи для слабослышащих абонентов в цифровых телекоммуникационных сетях	21-24
1.3	Анализ способов уменьшения влияния шума окружающей среды на разборчивость звуковых сигналов при использовании мобильных устройств телекоммуникации	25-35
1.4	Постановка задачи исследования	36-39
Выводы по главе 1		38-39
Глава 2.	Разработка способов и устройств, обеспечивающих клиенто-ориентированный доступ к ресурсам сетей, систем и устройств телекоммуникаций	40-94
2.1	Описание разработанного способа повышения разборчивости и понятности звуковых сигналов разного типа сложности в шумовой обстановке	40-56
2.2	Разработка способа компенсации потери слуха с помощью мобильных устройств	56-79
2.3	Прикладная реализация устройств на основе предложенных методов	80-93
Выводы по главе 2		93-96
Глава 3.	Разработка системы автоматического распознавания устной речи для мобильных устройств	97-127

3.1	Исследование существующих проблем применения систем автоматического распознавания устной речи в мобильных устройствах	96-106
3.2	Описание разработанного способа формирования базы данных сбалансированных синтагм и разработка алгоритма взаимодействия с БД	107-112
3.3	Разработка системы транскрибирования устной речи с использованием базы данных сбалансированных синтагм	112-125
Выводы по главе 3		125-127
Глава 4.	Исследование эффективности предложенных технических решений	128-140
4.1	Исследование эффективности разработанного способа и устройства повышения разборчивости полезного сигнала в агрессивной шумовой обстановке	128-132
4.2	Исследование эффективности разработанного способа и устройства формирования персонализированных сигналов для компенсации потери слуха с помощью ресурсов мобильных сетей и устройств связи	132-134
4.3	Исследование эффективности функционирования реализованного устройства на основе способа формирования базы данных сбалансированных синтагм	135-139
Выводы по главе 4		139-140
ЗАКЛЮЧЕНИЕ		141-143
СПИСОК ЛИТЕРАТУРЫ		144-153
Приложение А. Пример интерфейса разработанного программного обеспечения для сбора данных при проведении исследования эффективности (Мобильное приложение Hearing Aid)		154-155
Приложение Б. Копии актов внедрения		156-158

ВВЕДЕНИЕ. Актуальность темы исследования. Современные реалии таковы, что смартфоны и коммуникации через системы WiFi становятся повсеместны. По данным ежегодного отчета Digital 2024[1], в котором анализируются основные изменения цифрового пространства как на мировом рынке в целом, так и в каждой конкретной стране, на январь 2024 г.:

- число пользователей мобильных телефонов составило 5,61 миллиарда, что составляет 69,4% от общей численности населения. Кроме того, за последний год этот показатель увеличился на 138 миллионов (+2,5%);
- более 66% мирового населения (5,35 млрд) используют интернет в 2024 году. За последний год число пользователей и увеличилось на 97 млн.;
- в России было зарегистрировано 219,8 млн мобильных абонентских подключений, что превышает общее население страны - процент мобильных подключений к населению составил 152,5%.

Получение доступа к ресурсам сетей, систем и устройств телекоммуникации, таких как мобильная телефония, играет серьезную социально-экономическую роль, позволяя повысить качество жизни и вовлеченность людей с пониженным уровнем слуха в реалии современной жизни. Потеря слуха, даже частичная, существенно сказывается как на психическом и физическом здоровье человека, так и на социальной адаптации людей и их экономических возможностях.

По данным исследования аналитического бюро информационного портала GxP News, в России насчитывается **не менее 19 млн человек с проблемами слуха**, поэтому тема диссертации, направленная на повышения разборчивости и распознавания звуковых сигналов для доступа слабослышащих абонентов к ресурсам сетей, систем и устройств телекоммуникаций, актуальна.

Степень разработанности темы. Развитие технологий, обеспечивающих доступ к ресурсам цифровой экономики, людям с ограниченными возможностями (ассистивные технологии – theassistivetechologies), является важной экономической, социальной и гуманитарной задачей

любогосовременного общества, позволяя повышать уровень и качество жизни миллионам людей.

По данным Агентства инноваций г. Москвы в области развития ассистивных технологий в России разработка новейших технологий, связанных с развитием слуховых аппаратов равна нулю, как показано на рис. 1 [2].



Рисунок 1. Доля развития новейших технологий в области ассистивных технологий поддержки слабослышащих людей

Научный задел для повышения разборчивости и распознавания звуковых сигналов слабослышащими абонентами имеется, поскольку теория передачи звуковых сигналов успешно развивалась многими российскими и зарубежными учеными, такими как: Н.А. Гарбузов, А.И. Парфентьев, Ю.И. Бобровницкий, Н.Н. Андреев, Л.М. Бреховских, Е.Я. Юдин, Ю.М. Сухаревский, С.Н. Ржевкин, В.И. Шоров, О.А. Сапожников, И.Б. Есипов, И.Е. Кузнецова, С.Н. Куличков, G. Helmholtz, S. Naykin, P. Nil, M. Pankin и др. J. R. Rutkowski, J. H. L. Hansen, S. F. Quatieri и др.

Несмотря на известные достижения акустиков проблемы повышения разборчивости и распознавания звуковых сигналов слабослышащими абонентами пока решены недостаточно.

Объектом исследования являются ассистивные технологии для слабослышащих абонентов телекоммуникационных сетей

Предмет исследования – способы повышения разборчивости и распознавания речи, передаваемой в агрессивных акустических условиях окружающей среды.

Целью исследования является разработка комплексной технологии и устройств, обеспечивающих клиенто-ориентированный доступ слабослышащих абонентов к ресурсам сетей, систем и устройств телекоммуникаций.

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

1. Провести анализ методов повышения разборчивости звуковых сигналов в мобильных устройствах и цифровых сетях в условиях агрессивного окружающего шума.

2. Разработать метод повышения разборчивости и информативности речи в шумовой обстановке для обеспечения доступа слабослышащих абонентов к ресурсам сетей, систем и устройств телекоммуникаций.

3. Разработать способ компенсации потери слуха в телефонной системе в условиях агрессивной шумовой обстановки.

4. Реализовать устройство компенсации потери слуха на основе разработанного способа, позволяющее совместить функционалы телефонного и слухового аппарата.

5. Разработать программное обеспечение, позволяющее осуществлять дистанционную настройку устройства компенсации потери слуха в телефонной системе и автоматизацию процесса удаленного тестирования и расшифровки аудиограммы.

6. Провести анализ структуры, функционала и используемых методов в современных системах автоматизированного распознавания речи.

7. Разработать способ и реализовать программно-аппаратное обеспечение устройства автоматизированного распознавания речи на основе предложенного способа автоматизированного распознавания речи посредством текстового процессора с применением транскрипционного моделирования.

8. Оценить эффективность разработанных методов и средств повышения разборчивости и распознавания звуковых сигналов с помощью мобильных устройств для слабослышащих абонентов.

Научная и практическая новизна диссертации определяется следующим:

- Разработан и обоснован комплексный подход к реализации ассистивной технологии для обеспечения слабослышащим абонентам качественного доступа к ресурсам цифровых сетей, систем и устройств телекоммуникации.

- Предложен способ повышения разборчивости, информативности и натуральности звучания звуковых сигналов, в том числе и музыкальных композиций при прослушивании в акустической шумовой обстановке (Патент RU 2589298 C1 [3]).

- Разработан алгоритм работы блока обработки звуковых сигналов, реализующий предложенный метод [3].

- Создан блок обработки звуковых сигналов с адаптивным компрессором динамического диапазона, которым изменяют динамический диапазон полезного сигнала. [3].

- Разработан способ компенсации потери слуха путем формирования персонализированных аудиосигналов для слабослышащих пользователей на основе их атрибутов, хранящихся в базе данных на сервере сети связи и имеющих привязку к телефонным номерам. (Патенты RU 2568281 C2 [4], WO/2014/193264 [5], WO/2008/130265 [6], патент Китая 101558407 [7]).

- Разработан способ формирования сбалансированных синтагм, обеспечивающий уменьшение требуемых вычислительных ресурсов для автоматизированного преобразования речи в текст при сохранении уровня

точности распознавания устной речи 90% и увеличении скорости обработки на 25%. (Патенты RU 2460154 C1 [9], RU 2591640 C1[10], RU 2510954 C2[11], RU 2427044 C1 [12], WO/2013/180600 [13], RU 2393548 C1 [14]).

- Предложен алгоритм функционирования устройства автоматического распознавания речи на основе текстового процессора [9].

- Создана комплексная технология клиенто-ориентированного доступа слабослышащих абонентов к ресурсам цифровых сетей и систем.

Теоретическая значимость диссертационной работы заключается в разработке и исследовании новых способов:

- использования частотно-зависимого адаптивного усиления для повышения разборчивости аудио сигналов в системах связи и мультимедиа в условиях агрессивной окружающей среды (патент RU 2589298 C1[3]);

- формирования персонализированных аудиосигналов для слабослышащих пользователей на основе их атрибутов, полученных из аудиограмм - частотных характеристик слуха слабослышащего пользователя, хранящихся в базе данных на сервере сети связи и имеющих привязку к телефонным номерам слабослышащих пользователей (патент РФ RU 2568281 C22015[4-8, 17-24]);

- автоматизированного распознавания и преобразования устной речи в текст с учетом модификации и текстонезависимой конверсии голоса с помощью разработанного текстового процессора в режиме реального времени (патент РФ RU 2460154 C1 2012[9-16]).

Практическая значимость работы состоит в следующем:

1. Создание новых программно-аппаратных устройств, позволяющих:

- совмещать функции мобильного телефона с функцией слухового аппарата, уменьшая влияние шума окружающей среды и повышая разборчивость и естественность звуковых сигналов;

- увеличивать натуральность, разборчивость и понятность речи для конкретного слабослышащего пользователя при использовании мобильного телефона за счет формирования на сервере сети связи базы данных

персонализированных аудиосигналов на основе атрибутов конкретного абонента;

- в дистанционной форме настраивать разработанные устройства с учетом индивидуальных особенностей слабослышащего абонента и самостоятельно выполнять тестирование слуха;

- уменьшать требуемое количество вычислительных и информационных ресурсов в среднем на 25% при обеспечении достаточного уровня точности распознавания и преобразования речи в текст.

2. Реализовано устройство, позволяющее использовать мобильный телефон в четырех режимах:

а) слухового аппарата для прослушивания слабослышащим пользователем мультимедийных устройств,

б) слухового аппарата для разговора с человеком-собеседником, находящимся рядом со слабослышащим пользователем,

в) телефонного аппарата для слабослышащих [4-8].

г) системы автоматизированного распознавания и преобразования речи в текст в режиме реального времени с выводом транскрибированной речи на экран мобильного устройства.

3. Разработаны программные модули динамической компрессии аудиосигнала для инсталляции в мобильный телефонный аппарат слабослышащего пользователя [4].

4. Разработан комплекс программного обеспечения для мобильных телефонов, обеспечивающий дистанционную настройку телефона для работы в трех режимах и диагностирование слуха слабослышащих абонентов [25-31].

5. Реализовано программно-аппаратное обеспечение для функционирования устройства автоматизированной обработки текста (патент RU 2393548 C1[33], WO/2008/118038 [34, 9-14]).

6. Разработаны программные модули динамической компрессии аудиосигнала для инсталляции в мобильный телефонный аппарат слабослышащего пользователя [4].

7. Разработан комплект программного обеспечения для мобильных телефонов, обеспечивающий дистанционную настройку телефона для работы в трех режимах и диагностирование слуха слабослышащих абонентов [25-31].

8. Реализовано программно-аппаратное обеспечение для функционирования устройства автоматизированной обработки текста ([9-16], [32-34]).

Реализация результатов работы. Разработанные методы и полученные результаты использованы при проектировании и реализации мобильных приложений и устройств:

- «Слуховой аппарат для мобильной платформы», предназначенное пользователям с нарушениями слуха компанией ООО «Алгоритм Электроникс»;
- для коррекции слуха, предназначенного для использования на смартфонах под управлением iOS и Android компанией Dialog Axiata PLC (Шри-Ланка),
- для дистанционного обучения слабослышащих студентов компанией образовательного частного учреждения дополнительного профессионального образования «Бауманский компьютерный учебный центр «Специалист»».

Методы исследования. В работе использованы положения теории электроакустики, теории информации, теории помехоустойчивого кодирования, методы искусственного интеллекта и машинного обучения, методы транскрипционного моделирования.

Основные положения, выносимые на защиту:

1. Способ повышения разборчивости, информативности и натуральности звучания звуковых сигналов, в том числе и музыкальных композиций при прослушивании в акустической шумовой обстановке. (Патент RU 2589298 C1, [3]).

2. Способ формирования персонализированных аудиосигналов для слабослышащих пользователей на основе их атрибутов, полученных из аудиограмм - частотных характеристик слуха слабослышащего пользователя, хранящихся в базе данных на сервере сети связи и имеющих привязку к телефонным номерам слабослышащих пользователей.

3. Модифицированный способ автоматического распознавания, синтеза и преобразования речи в текст, обеспечивающий уменьшение используемых вычислительных и информационных ресурсов в среднем на 25% при обеспечении достаточного уровня точности.

4. Аппаратное и программное обеспечение устройства формирования персонализированных аудиосигналов для слабослышащих пользователей на основе их частотных характеристик слуха, обеспечивающее натуральность, разборчивость и понятность речи в условия повышенного шума окружающей среды.

5. Аппаратное и программное обеспечение устройства автоматической транскрипции устной речи, позволяющее уменьшить требуемые информационные ресурсы на 25% при использовании в мобильных телефонах.

Степень достоверности и апробация результатов. Достоверность результатов исследования обусловлена корректностью математических выкладок, адекватностью моделей применительно к системам автоматизированного распознавания речи, согласованием результатов теоретического анализа с данными экспериментальных исследований, положительными результатами внедрения.

Результаты работы обсуждались на международных конференциях:

- XVI Международная научная конференция «Физика и радиоэлектроника в медицине и экологии» с научной молодежной школой им. И.Н. Спиридонова ФРЭМЭ 2024» 02-04 июля 2024;

- VIII Международная научно-практическая конференция «Актуальные проблемы и перспективы развития радиотехнических и инфокоммуникационных систем» «РАДИОИНФОКОМ-2024»

- LXIII International Scientific-Practical Conference «EurasiaScience». – г. Москва. – 2024 г.

- XXVII Международная научно-практическая конференция «Наука и технологии: модернизация, инновации, прогресс. – г. Анапа. – 2024 г.

- IEEE AEIT Annual 2015 Conference. Naples, Italy -2015 г.

Публикации. Основные результаты диссертации изложены в 33 опубликованной работе, в том числе в 4 статьях в журналах из перечня ВАК, в 1 статье в изданиях, индексируемых в базе WoS и Scopus, в 25 патентах по теме работы и в 3-х других изданиях.

Соответствие паспорту специальности. Содержание диссертации соответствует пунктам 1, 2, 6, 18 паспорта научной специальности 2.2.15 - «Системы, сети и устройства телекоммуникаций».

Структура и объем диссертации. Диссертация состоит из введения, четырех разделов, заключения и приложения. Общий объем диссертационной работы вместе с приложениями составляет 158 страниц, в том числе 141 страниц основного текста, 42 рисунков, 10 таблиц, списка литературы из 109 наименований и 2-х приложений.

Личный вклад автора. Изложенные в диссертации теоретические и экспериментальные результаты исследования получены соискателем лично, либо при его непосредственном участии. Из 25 патентов по теме работы 3 патента получены без соавторства –RU Patent RU 2510954 C2 [11]; Patent WO/2013/180,600 [13]; Patent MX/a/2009/011,194 [21].

ГЛАВА 1. ИССЛЕДОВАНИЕ ПРОБЛЕМ РАЗБОРЧИВОСТИ ЗВУКОВЫХ СИГНАЛОВ ДЛЯ ДОСТУПА СЛАБОСЛЫШАЩИХ АБОНЕНТОВ К РЕСУРСАМ СЕТЕЙ, СИСТЕМ И УСТРОЙСТВ ТЕЛЕКОММУНИКАЦИЙ.

1.1. ТЕКУЩЕЕ СОСТОЯНИЕ ПРОБЛЕМЫ РАЗБОРЧИВОСТИ ЗВУКОВЫХ СИГНАЛОВ ДЛЯ СЛАБОСЛЫШАЩИХ АБОНЕНТОВ

Доступность людей с нарушениями слуха к нормальной активной социальной жизни в настоящее время ограничена не только в развивающихся, но и в ряде развитых стран. По данным Всемирной организации здравоохранения (ВОЗ) за 2023 г., около 430 млн человек в мире необходима реабилитационная помощь для решения проблемы инвалидизирующей потери слуха (потеря слуха в слышащем лучше ухе от 35 дБ). Эти нарушения могут быть спровоцированы рядом причин на разных этапах жизни человека. На пренатальном периоде жизни влияние на возникновение нарушений оказывают генетические факторы и внутриутробные инфекции, на более поздних этапах нарушения слуха вызываются менингитом и хроническими отитами, а также громкими шумами.

Самой распространенной формой снижения слуха является нейросенсорная тугоухость, при которой поражаются участки звуковоспринимающего отдела слухового анализатора. В частности, у пожилых людей самым распространенным нарушением слуха является двусторонняя симметричная возрастная нейросенсорная тугоухость (пресбиакузис или возрастная потеря слуха), которая возникает вследствие дегенерации улитки внутреннего уха или связанных с ней структур. Треть пожилого населения планеты (65 лет и старше) – слабослышащие, согласно данным ВОЗ [35]. На рисунке 2 показано соотношение населения с нарушениями слуха по странам. В частности, в США около 14,8% населения имеют нарушения слуха [36], в Китае

18% населения имеют нарушения слуха, в России 9% населения имеют нарушения слуха [37].

Полная и частичная потеря слуха— это снижение или потеря слуха в результате поражения звуковоспринимающего аппарата. В органе слуха выделяют звукопроводящий и звуковоспринимающий отделы. Звукопроводящий отдел предназначен для передачи звуковых колебаний к рецептору, можно интерпретировать его как приемник сигнала, звуковоспринимающий отдел отвечает за реакцию нервной ткани на звуковое раздражение – обработка сигнала.

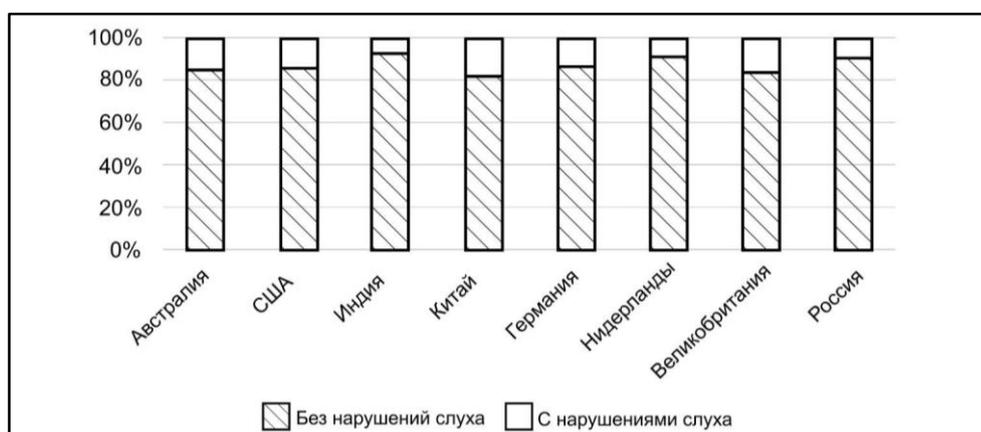


Рисунок 2. Соотношение населения стран с нарушениями слуха

Нарушение слуха означает снижение чувствительности слуха и снижение разборчивости речи. Чувствительность тонального слуха, или порог слухового восприятия, означает восприятие человеком самого тихого звука определённой частоты (тона) при проведении тональной аудиометрии. Техническое решение проблемы на сегодняшний день – слухопротезирование, то есть повышение качества восприятия информации с помощью дополнительных технических средств – слуховых аппаратов.

Современный слуховой аппарат (СА) – это сложное цифровое устройства - электроакустический прибор, усиливающий поступающий извне звуковой сигнал и преобразующий звуковые волны в электрические сигналы. В СА

применяют два типа микрофонов: разнонаправленные - улавливающие все окружающие звуки и подходящие для спокойной обстановки, и направленные, принимающие звуки, приходящие лишь из определенного направления и позволяющие общаться с конкретным собеседником в сложной акустической ситуации. Миниатюризация всех компонентов позволяет создавать небольшие слуховые аппараты – от носимых (карманных) до внутриушных и внутриканальных.

Характеристики СА подбираются индивидуально в зависимости от характера снижения слуха пользователя. Для определения требуемых параметров СА необходимо выполнить диагностику слуха, в том числе методом аудиометрии, при которой в звукоизолированных помещениях с использованием специального оборудования пользователю воспроизводятся тональные сигналы определенных частот и фиксируется порог восприятия каждого сигнала[33]. Полученный результат аудиометрии формируется в виде графика - аудиограммы для каждого уха, где каждой частоте соответствует значение порога слышимости. В зависимости от величины падения слуха различают степени тугоухости от I до IV.

Современные слуховые аппараты представляют собой сложные акустические цифровые программируемые устройства, в которых происходит индивидуальная обработка звуков по нескольким частотным диапазонам в соответствии с заданным алгоритмом, настраиваемая специалистом – аудиологом по параметрам снижения слуха пользователя.

Все специализированные устройства обладают примерно одинаковым функционалом и построены на аналогичных принципах [38-39]. Принципиальный состав современного слухового аппарата: микрофон, цифровой процессор, усилитель-преобразователь и динамик (рисунок 3).

Микрофон преобразовывает электроакустические волны в электрические колебания, процессор анализирует полученные звуки и обрабатывает их. После чего обработанный звук перемещается в усилитель, который по уникальному

алгоритму усиливает входящий сигнал и позволяет настраивать необходимые параметры СА с учетом индивидуальных особенностей пациента с помощью процессора, а динамик преобразовывает выходной сигнал усилителя обратно в акустический.

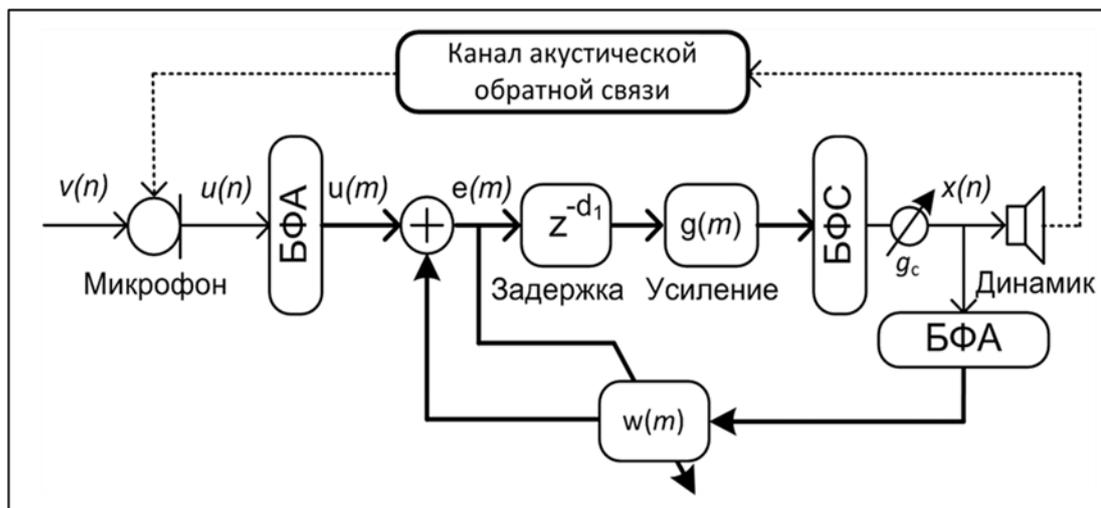


Рисунок 3. Принципиальная схема слухового аппарата, где БФА – банк фильтра анализа, БФС – банк фильтра синтеза

Основным функционалом современных СА является:

- частотно-зависимое усиление;
- компрессия;
- фильтрация и шумоподавление;
- подавление акустической обратной связи.

Аудиограмма позволяет выполнить расчет необходимого (целевого) усиления и выходной мощности, обеспечивающих максимально комфортное для пользователя звучание. Акустическим усилением называется разность между выходным уровнем звукового давления, создаваемого слуховым аппаратом в имитаторе уха и входным уровнем звукового давления, измеренным в тестовой точке. Величина максимального акустического усиления определяет применимость СА к различным степеням потери слуха.

Компрессия – процесс выравнивания громкости звука в аппарате таким образом, чтобы тихие звуки усиливались в соответствии с заданными

параметрами, а громкие оставались на прежнем уровне (без усиления). Таким образом слух предохраняется от усиления громких звуков до неприемлемых значений. Алгоритмы компрессии реализуются только в топовых моделях слуховых аппаратов.

Одним из недостатков компактных и миниатюрных СА является возникающая в них акустическая обратная связь – прием микрофоном аппарата и усиление звука, исходящего из динамика самого аппарата. Для подавления этого эффекта в слуховых аппаратах используются специальные фильтры и алгоритмы подавления эхосигнала, усложняющие процесс обработки звука.

Важными параметрами любого СА являются частота дискретизации и задержка обработки сигнала. Частота дискретизации определяет полосу пропускания аппарата, чем выше частота дискретизации – тем больший диапазон звуков доступен пользователю. В современных цифровых СА полоса пропускания составляет около 10 кГц, так как повышение частоты дискретизации, во-первых, требует больших вычислительных ресурсов и приводит к повышенному энергопотреблению и, во-вторых, из-за того, что при наиболее распространенном типе потери слуха страдает в основном восприимчивость высоких частот, а разработчикам СА всегда приходится выбирать - увеличивать полосу пропускания или экономить ресурс батарей слухового аппарата.

Задержка обработки сигнала складывается из продолжительности всех этапов обработки звука в СА. Слишком высокая временная задержка приводит к окрашиванию звука собственного голоса. Таким образом, чем меньше задержка обработки сигнала, тем комфортнее звук аппарата для пользователя.

Для обработки используются банки фильтров анализа (БФА) и синтеза (БФС), производится подавление акустической обратной связи, при этом параметры усиления и компрессии подбираются и настраиваются по результатам аудиометрии пользователя (рисунок 3).

Устройства с более высоким технологическим уровнем позволяют улавливать звуки более натурально. Поэтому многие СА отличаются качеством звучания, индивидуальными параметрами, возможностью локализации звуков, использованием технологии адаптивного шумоподавления, подавления обратной связи, адаптивной направленности, позволяющие увеличить разборчивость речи.

Существующие в мире системы страхования, которые покрывают затраты пациента на покупку слуховых аппаратов, требуют получения консультации врача (сурдолога или отоларинголога) или специалиста-аудиолога и прохождения тестирования слуха для получения аудиограммы. Тестирование методом тональной аудиометрии проводится в помещении, которое соответствует международному стандарту ГОСТ Р ИСО 8253-1-2012 [40]. Такое аудио-логическое исследование, проверка и настройка корректирующих устройств требуют привлечения специалистов-аудиологов и специализированного оборудования. В результате затруднено получение помощи для населения регионов со слаборазвитой медицинской и транспортной инфраструктурой. В свою очередь, низкая доступность слуховых аппаратов обусловлена их высокой ценой, затратами на поддержание в работоспособном состоянии и спецификой первоначальной настройки.

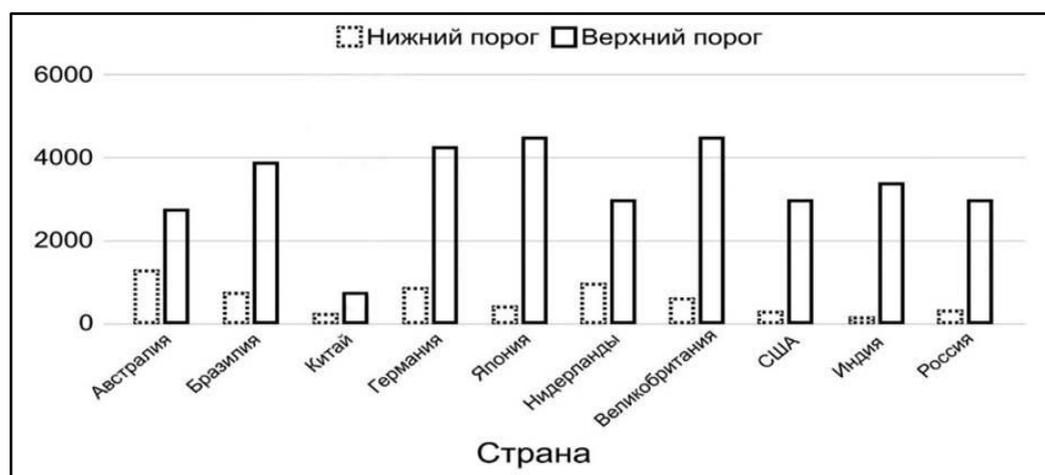


Рисунок 4. Стоимость слуховых аппаратов

Слуховые аппараты (СА) - узкоспециализированные устройства и этот фактор определяет высокую себестоимость производства и конечную цену для потребителя. Ценовой диапазон слуховых аппаратов на рынке Китая составляет \$230-670. Диапазон цен СА на рынке США - \$500-3000, на рынке Бразилии - \$780-3900 [41] (рисунок 4). В случае слухопротезирования обеих ушей расходы на слуховые аппараты удваиваются.

Технические характеристики слуховых аппаратов также значительно зависят от разных ценовых категорий. Практически любой слуховой аппарат способен существенно улучшить способность воспринимать звук в тихой окружающей обстановке. А вот по отношению сигнал/шум слуховые аппараты отличаются, и порой очень существенно. Функция шумоподавления, т.е. способность «очищать» человеческую речь от фонового шума может быть недостаточно производительной или вовсе отсутствовать. На рисунке 5-а изображены диапазоны восприятия сигнала человеком с нормальным слухом, а на рисунке 5-б - человеком с нейросенсорной тугоухостью.

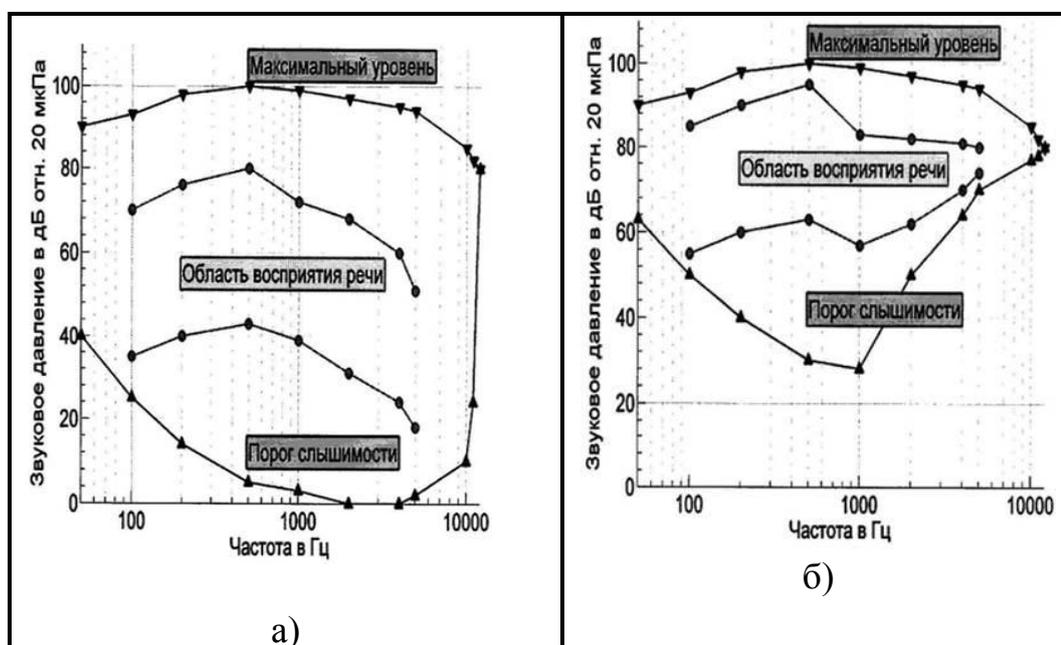


Рисунок 5. Диапазон восприятия сигнала

а) человека с нормальным слухом; б) человека с поврежденной слуховой системой

Задача, которую решают современные цифровые слуховые аппараты, это - преобразование характеристики для слабослышащего пользователя в характеристику нормально слышащего человека. Основной сложностью, возникающей при проектировании слуховых аппаратов, является ограничение допустимой задержки, вносимой в звуковой сигнал. При наличии большой задержки (свыше 8 мс) возникает паразитное эхо, которое отрицательно сказывается на восприятии. В современных СА обработка выполняется в частотных субполосах сигнала, что требует применения банков фильтров анализа и синтеза, которые вносят дополнительную групповую задержку и не обеспечивают задержки ниже 6-8 мс. Реализация доступа к ресурсам сетей, систем и устройств телекоммуникации при использовании слуховых аппаратов также сильно отличается в зависимости от ценовой категории, но в любом случае резко понижает комфортность коммуникации за счет уменьшения естественности передаваемых звуковых сигналов, понятности и разборчивости речи в условиях повышенного шума.

В диссертационной работе предлагается решить данную проблему с помощью возможностей современных сетей, систем и устройств телекоммуникации. Современными смартфонами могут стать эффективным решением для диагностики и компенсации нарушений слуха при оснащении их специальным аппаратным и программным обеспечением.

1.2. АНАЛИЗ СПОСОБОВ ПОВЫШЕНИЯ РАЗБОРЧИВОСТИ РЕЧИ ДЛЯ СЛАБОСЛЫШАЩИХ АБОНЕНТОВ В ЦИФРОВЫХ ТЕЛЕКОММУНИКАЦИОННЫХ СИСТЕМАХ

Люди с нейросенсорной тугоухостью обычно имеют приподнятый порог восприятия акустической информации, который мешает услышать звуки низкой интенсивности. Однако восприятие громких звуков часто такое же, как и у здоровых людей. Порог потери слухового восприятия является частотно зависимым и определяется для каждого пациента на заданных частотах (200, 500, 1000, 2000, 3000, 4000 и 6000 Гц), используя чистые тональные сигналы.

Задача повышения разборчивости у тугоухих людей заключается в компоновке динамического диапазона речевых и повседневных звуков в ограниченный динамический диапазон нарушенного слуха. Этот процесс компрессии динамического диапазона представляет собой отображение слышимого диапазона сигнала в область остаточного восприятия пациента. При этом усиленный сигнал не должен превышать максимального уровня, в противном случае это вызовет болевые ощущения у человека. Более того, тугоухость обычно частотно зависимая, т.е. компрессор должен выдерживать различные динамические уровни в различных частотных диапазонах. Решение данной проблемы в общем случае достигается применением многоканальных систем, таких как банки фильтров с различной степенью компрессии в каждом канале. При проектировании многоканального компрессора динамического диапазона требуется найти:

а) компромисс между частотным разрешением и временной задержкой. У известных решений увеличение частотного разрешения анализа приводит к увеличению времени обработки сигнала;

б) максимально согласовать частотное разрешение многоканального компрессора динамического диапазона с частотным разрешением восприятия акустической информации человеком;

в) компромисс между частотным разрешением и групповой задержкой.

Задержка обработки сигнала в многоканальном компрессоре динамического диапазона с неравнополосным банком фильтров больше в диапазоне нижних частот, чем в высокочастотных каналах, однако она меньше, чем в равнополосных системах с аналогичным частотным разрешением. При наличии большой задержки (свыше 8 мс) возникает паразитное эхо, которое отрицательно сказывается на восприятии и ухудшается разборчивость.

Известен способ повышения разборчивости речи у пользователей с нейросенсорной тугоухостью в цифровых телекоммуникационных системах, базирующийся на слуховом аппарате (СА) и мобильном телефонном аппарате (ТА), когда для приема сигнала из ТА, пользователь подносит его к своему СА [42]. Но в этом способе возникают проблемы как акустической совместимости, так и электромагнитной совместимости данных двух приборов.

Обеспечение акустической совместимости обуславливает необходимость изменения усиления в динамиках ТА и чувствительности микрофона в СА, что может приводить к образованию обратной акустической связи в системе, и, как следствие, понижение разборчивости речи, а при определенных уровнях усиления и чувствительности микрофона к болезненным ощущениям пользователя. Например, в США, разработан специальный стандарт на проверку совместимости разных типов СА и мобильных ТА между собой [43].

В настоящее время часто добавляется дополнительное устройство к СА, имеющее автономное питание и работающее с ТА через беспроводной канал, например, Bluetooth [44]. Это позволяет разнести СА и ТА на определенное расстояние друг от друга, при этом исключив проблему акустической и электромагнитной совместимости двух приборов. Однако недостатком такого способа является необходимость дополнительного устройства - ретранслятора сигнала между СА и ТА.

Известен также способ повышения разборчивости речи у слабослышащих пользователей в цифровых телекоммуникационных системах, базирующийся на СА с встроенным беспроводным каналом связи [45]. Этот способ включает

формирование входного аудиосигнала путем микширования сигнала микрофона с аудиосигналом, принятым из канала беспроводной связи от удаленного терминала (телевизор, мультимедийный проигрыватель, другой СА с беспроводным каналом связи и других устройств аудиосигналов со встроенными беспроводными каналами связи), динамическую компрессию входного аудиосигнала, заключающуюся в формировании субполосных аудиосигналов и управлении уровнем сигнала в соответствующих субполосах для обеспечения требуемой динамики уровней субполосных сигналов, обусловленных аудиограммой пользователя СА, последующее восстановление аудиосигнала на основе банка фильтров синтеза. Достоинством данного аналога со встроенным беспроводным каналом связи является исключение негативного влияния шумовой обстановки и реверберации в помещении на разборчивость речи пользователем СА в принятом аудиосигнале от удаленного терминала.

Имеется возможность подключить пользователей в сеть СА, которая не подвержена влиянию акустической обстановки в помещении, и транслировать, например, экстренные сообщения. Недостатком здесь является повышенная сложность и стоимость СА, короткое расстояние связи, увеличенная потребляемая мощность СА. Но главным ограничением данного способа является то, что организация общения пользователей СА через ТА предполагает использование по существу первого способа со всеми присущими ему недостатками, проблемой как акустической совместимости, так и электромагнитной совместимости СА и ТА.

Реализовано на сегодняшний день персональное коммуникационное устройство, включающее передатчик и приемник, которые передают и принимают коммуникационные сигналы, кодирующие аудио сигналы, аудио-преобразователь, который делает аудиосигнал слышимым, микрофон, и схему управления, связанную с передатчиком, приемником, аудио- преобразователем и микрофоном, содержащую логику, применяющую многополосную компрессию к аудиосигналам, включая генерацию параметров для

многополосной компрессии на основе хранимых пользовательских данных и на основе данных окружающей среды, установленные при управлении преобразователем, делающим аудиосигналы слышимыми. Устройство, способно содержать три подпрофиля - аудиофиль пользователя, хранимый и получаемый, в том числе, с помощью удаленных средств, обладающих возможностью осуществить передачу аудиофиль устройству, профиль персональных предпочтений пользователя, данные об окружающей среде, а именно профиль окружающего шума, которые в любой комбинации друг с другом применяются к аудиосигналу, получаемому после декодирования коммуникационного сигнала, поступающего на приемник устройства.

Ограничениями этого устройства являются:

- невозможность использования пользователем данного устройства персонализированного аудиосигнала, необходимого во время личного общения с человеком-собеседником;
- отсутствие возможности пользователем данного устройства получить персонализированный аудиосигнал при одновременном общении пользователя с другим абонентом по телефону и с собеседником лично, при приеме аудиосигнала из динамиков различных устройств, например во время просмотра передач телевидения, прослушивания музыки и т.п.;
- неосуществимость получения персонализированного аудиосигнала слабослышащими абонентами сетей связи, не имеющими подобных устройств, непосредственно из сетей связи, в частности при общении пользователя известного устройства с другим слабослышащим абонентом сети связи, не обладающего подобным устройством;
- отсутствие у пользователя данного устройства в соответствии с его предпочтениями реализации различных режимов: телефонного разговора, личного общения с собеседником, приема аудиосигнала из динамиков разнообразных устройств.

1.3. АНАЛИЗ СПОСОБОВ УМЕНЬШЕНИЯ ВЛИЯНИЯ ШУМА ОКРУЖАЮЩЕЙ СРЕДЫ НА РАЗБОРЧИВОСТЬ ЗВУКОВЫХ СИГНАЛОВ ПРИ ИСПОЛЬЗОВАНИИ МОБИЛЬНЫХ УСТРОЙСТВ ТЕЛЕКОММУНИКАЦИИ

Необходимо остановиться также на той проблеме, что при реализации коммуникации слабослышащий абонент окружен шумом, например разговаривающими людьми, транспортными шумами и т.п, чьи звуки будут маскировать голос телефонного собеседника и отрицательно влиять на разборчивость его речи.

Для повышения разборчивости используют следующие методы [46-49]:

- **снижение уровней помех**, которое достигается за счет приближение микрофона или увеличение уровня голоса говорящего, что в условиях коммуникации слабослышащего человека с помощью мобильного телефона практически невозможно реализовать. Существуют алгоритмы цифрового устранения стационарных шумов, также появились фильтры, которые построены на обученных нейросетях и позволяют вычитать «деаэрация» разные типы шумов, типа стук клавиш клавиатуры, лай собаки и т.д. Возможности искусственного интеллекта позволяют выделять определенный голос из множества голосов. Безусловно, при использовании нейронных сетей появляется большая (для реального масштаба времени) задержка в 100-200 мс, но для линии связи такая задержка не критична;

- **повышение уровня звукового давления у слушателя по прямому звуку**, которое достигается приближением микрофона к источнику звука, повышением уровня голоса говорящего, увеличением индекса тракта;

- **сжатие речевого сигнала D** — повышение уровней звукового давления слабых звуков при сохранении максимальных уровней звукового давления. Предельным случаем сжатия D является амплитудное ограничение, так называемое **клиппирование**. В этом методе речевой сигнал превращается в последовательность импульсов постоянной амплитуды, но с меняющимися

интервалами между нулевыми переходами (телеграфный режим работы). В результате все звуки речи будут иметь одинаковый (максимальный) уровень при приеме. Качество звучания в этом случае ухудшается, но разборчивость резко повышается, так как слабые звуки неклипированной речи, маскирующиеся помехами, при таком способе передачи окажутся выше уровней помех. Такой способ также может использоваться крайне осторожно, так как в условиях максимального приближения слухового аппарата качество звучания влияет на комфортность восприятия речи и повышение разборчивости речи в ущерб качеству звучания не является возможным.

Сейчас получил распространение способ широкополосной динамической компрессии, (Wide Dynamic Range Compression - WDRC), которая автоматически регулирует уровень громкости входящего звукового сигнала, делая тихие звуки громче, а громкие — тише, чтобы обеспечить комфортное прослушивание [51].

Прослушивание различных звуковых сигналов (таких как музыкальные произведения или речевые сообщения) в шумовой обстановке (рисунок 6) часто бывает затруднительным, поскольку акустический шум $v[n]$ при соответствующей интенсивности маскирует полезный сигнал $x[n]$.

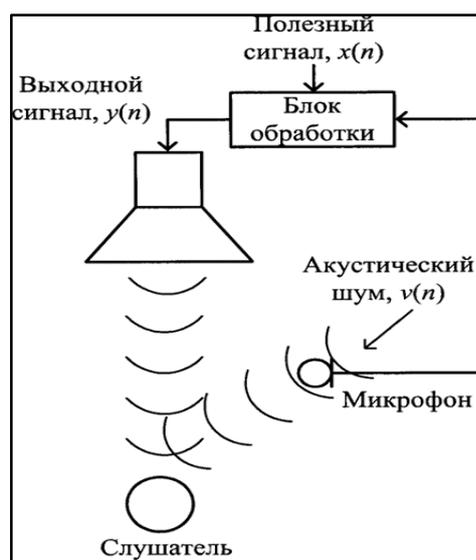


Рисунок 6. Схема восприятия акустической информации в шумовой обстановке

Под полезным сигналом понимается сигнал без акустического шума, который «Слушатель» воспринимает как выходной сигнал $y[n]$ блока обработки через громкоговоритель в обстановке агрессивного акустического шума.

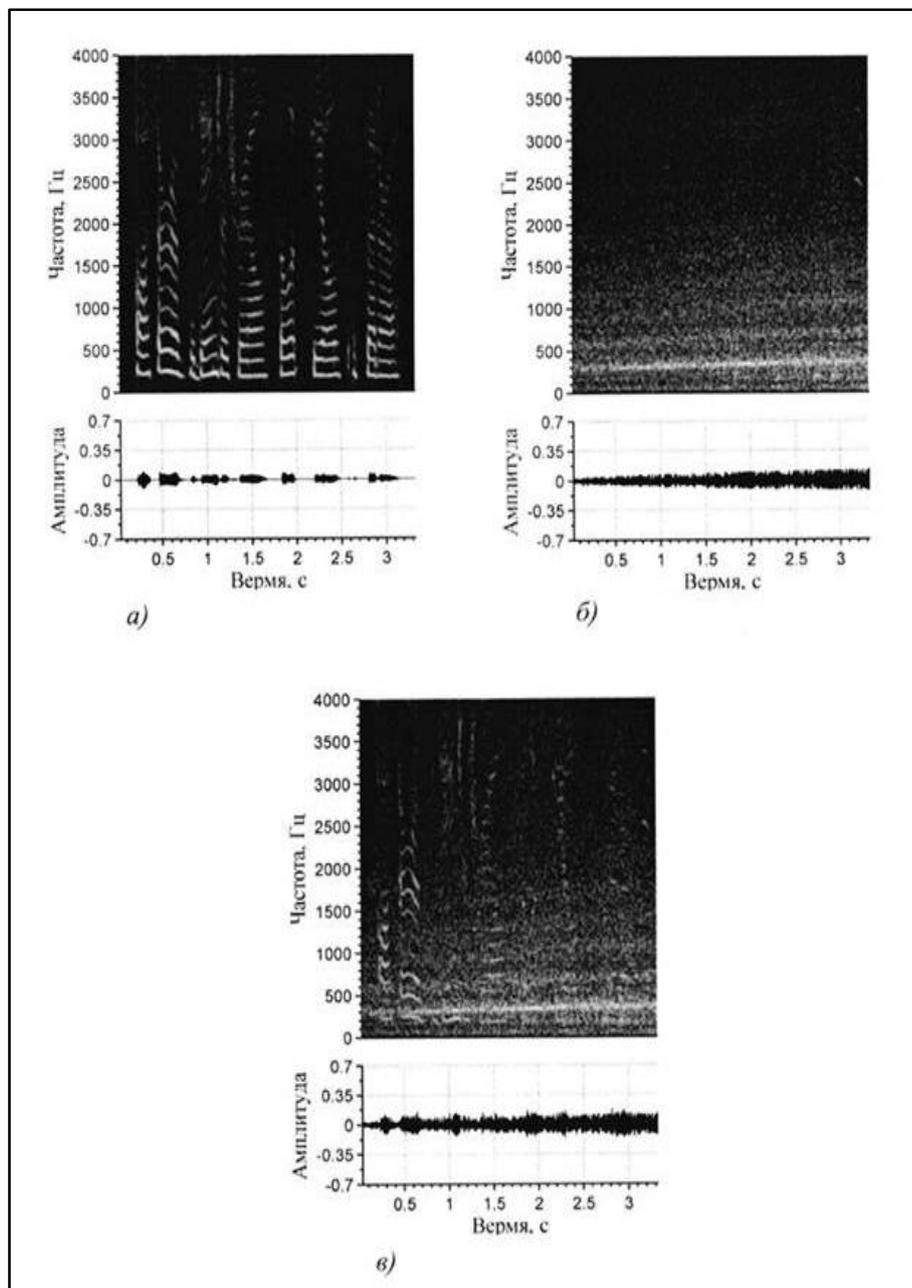


Рисунок 7. Частотно-временное представление сигнала (спектрограммы):
а) чистый речевой сигнал; б) сигнал шума метрополитена;
в) сумма сигналов речи и шума для отношения сигнал-шум - 5 дБ

В качестве полезного сигнала $x[n]$ может выступать сигнал мультимедийных воспроизводящих устройств, сигнал в системах видеоконференций, громкой связи, IP-телефонии и т.п. В системах связи

различного назначения полезный сигнал $x[n]$ - сигнал дальнего диктора, приходящий из дальнего диктора $x[n]$ затруднено акустическим шумом $v[n]$ окружающей сети, а «Слушатель» - ближний диктор, для которого восприятие сигнала обстановки, в которой он находится.

Схема восприятия акустической информации в шумовой обстановке и соответствующие спектрограммы проиллюстрированы на рисунках 7-а, 7-б, 7-в. Например, при покрытии чистого речевого сигнала шумом метрополитена этот речевой сигнал едва может быть идентифицирован (рисунок 7-в).

Проанализируем некоторые типовые методы обработки звуковых сигналов, применяющиеся на практике для уменьшения акустического шума [45, 46]:

- устранение эффекта слуховой маскировки полезного сигнала помехами и искажениями в частотной и временной областях;
- устранение фоновых шумов различных типов с целью снижения напряженности внимания и утомляемости «Слушателя» при прослушивании;
- допустимое для сохранения разборчивости сокращение полосы частот прослушиваемого зашумленного сигнала с целью снижения утомления «Слушателя», удаление низкочастотного гула и высокочастотного шипа;
- выравнивание больших пиков и перепадов амплитуды сигнала и громкости отдельных участков обрабатываемой исходной звукозаписи, снижение амплитуды сигнала в паузах без речи;
- удаление или снижение амплитуды импульсных помех и других интенсивных посторонних звуков;
- удаление регулярных медленно меняющихся помех (музыкальных, транспортных, производственных, сетевых, снижение эффекта реверберации);
- сглаживание и выравнивание спектра сигнала (амплитудно-частотной характеристики канала);
- дополнительная режекция (вырезание из спектра одной или нескольких узких полос частот) узкополосных помех;

- удаление аддитивных широкополосных шумов («шип» магнитофонной ленты, радиоканала, микрофона);
- формирование у выходного сигнала амплитудной и частотной характеристики, оптимальной именно для слуха конкретного «Слушателя»;
- компенсация существенных фазовых искажений сигнала.

Не все из перечисленных способов эффективны для определенного типа помех, рассмотрим это подробнее.

Обеспечение комфортного прослушивания и достаточной разборчивости сравнительно легко достигается путем акустического усиления полезного сигнала $x[n]$, что позволяет вывести его из-за порога маскирования шумом $v[n]$. Большинство звуковоспроизводящих устройств имеет возможность ручной регулировки громкости, которая позволяет слушателю адаптировать режим воспроизведения к различной шумовой обстановке. Однако в контексте повышения разборчивости ручное управление громкостью имеет следующие ограничения:

- ручная регулировка не всегда доступна для оперативной реакции на изменение шумовой обстановки;
- для сигналов, имеющих широкий динамический диапазон, имеющих тихие и громкие фрагменты, нельзя подобрать фиксированный коэффициент усиления, поскольку увеличение громкости, достаточное для прослушивания тихих звуков, приводит к тому, что громкие звуки становятся слишком громкими (возможно превышающими болевой порог);
- усиление выполняется одинаково для всех частотных составляющих сигнала без учета спектральной плотности мощности маскирующего шума.

Известные системы редактирования шума осуществляют фильтрацию шума сигнала микрофона ближнего диктора до передачи в сеть. Однако часто требуется повысить разборчивость сигнала дальнего диктора (полезного сигнала $x[n]$) для ближнего диктора, находящегося в шумной акустической обстановке, когда шум прослушивается без какой-либо возможности быть подавленным. В

1960-х и 1970-х годах был выполнен ряд исследований по этой теме [45]. Для повышения разборчивости речи у ближнего диктора-слушателя предлагалось выполнять высокочастотную фильтрацию чистого сигнала дальнего диктора (полезного сигнала $x[n]$), чтобы усилить более высокие форманты, сопровождаемую амплитудной компрессией, чтобы защититься от искажений белого шума и уменьшить мощность шумовой среды соответственно. Способ предназначен для обработки исключительно речевых сигналов и предполагает наличие в сигнале специфических для речи характеристик, таких как формантные частоты и вокализованные фрагменты. Этот способ не применим к обработке неречевых звуковых сигналов (например, музыкальных произведений).

Существует известное техническое решение задачи повышения разборчивости речи в акустических шумах, основанное на максимизации индекса разборчивости SII (Speech Intelligibility Index), в котором повышение индекса разборчивости SII достигается путем модификации спектра речевого сигнала - полезного сигнала $x[n]$ с учетом спектральной плотности мощности сигнала шума $v[n]$. Индекс разборчивости речи SII является стандартизированной объективной мерой, которая отражает действительную разборчивость речи для различных неблагоприятных условий прослушивания [56]. Данная мера разборчивости разработана исключительно для анализа речевых сигналов и не может в исходном виде применяться к произвольным звуковым (например, музыкальным записям), так как критерием разборчивости речи является способность идентифицировать отдельные слоги речевого сообщения. Музыкальное же произведение (сообщение) имеет неслоговую структуру и воспринимается слуховой системой при помощи совсем иных интерпретационных механизмов.

Известен способ повышения разборчивости речевого сигнала - полезного сигнала $x[n]$ в шумах с сохранением исходной энергии речевого сигнала [57]. Способ состоит из двух последовательных этапов, которые осуществляют

адаптивное изменение формы спектральной огибающей и расширение/компрессия динамического диапазона. Этот способ обработки основывается на результатах исследования влияния формантной структуры чистой (незашумленной) речи на разборчивость. Адаптивное изменение формы спектральной огибающей выполняется путем подчеркивания формантной структуры речи и уменьшения наклона спектральной огибающей корректирующим фильтром. Степень обработки зависит от уровня вокализованности речевого фрагмента - чем ближе он к гласному звуку, тем сильнее вносимый эффект. Данный способ повышения разборчивости речевого сигнала - полезного сигнала $x[n]$ в шумовой обстановке выполняется в блоке обработки (рисунок 6) и состоит из следующих шагов:

1. Вычисляется вероятность $P_v[n]$ (или степень) вокализованности текущего речевого фрейма - «полезного сигнала» $x[n]$,

$$P_v[n] = \alpha \frac{rms[n]}{z[n]}, \quad (1.1)$$

где $\alpha = 1/\max(P_v[n])$ - является нормирующим множителем,

$rms[n]$ - среднеквадратичное значение отсчетов сигнала,

$z[n]$ - число переходов через ноль,

n - номер отсчета, $n=1, 2 \dots N$.

Статистические оценки $rms[n]$ и $z[n]$ выполняются на одном фрейме речевого сигнала с центром в n -м отсчете сигнала и продолжительностью 8,3 мс либо 4,5 мс в зависимости от того, обрабатывается мужской голос или женский. Фреймы набираются с шагом 10 мс.

2. Каждый фрейм речевого сигнала $x[n]$ длиной N отсчетов умножается на окно Хэннинга и вычисляется дискретное преобразование Фурье. Используя амплитудный спектр, оценивается спектральная огибающая сигнала $\hat{E}[\omega_k]$, где: ω_k - угловая частота, k - индекс частотной составляющей.

Коррекция огибающей $\hat{E}[\omega_k]$ выполняется при помощи трех последовательных фильтров:

$$\hat{E}[\omega_k] = E[\omega_k] H_s[\omega_k] H_p[\omega_k] H_r[\omega_k], \quad (1.2)$$

причем:

$H_s[\omega_k]$ и $H_p[\omega_k]$ являются адаптивными фильтрами,

$H_r[\omega_k]$ - фильтром с постоянными параметрами.

Для получения частотной характеристики фильтра $H_s[\omega_k]$ вычисляется наклон спектральной огибающей $T[\omega_k]$ при помощи следующего выражения:

$$\log T[\omega_k] = c_0 + 2c_1 \cos(\omega_k), \quad (1.3)$$

$$c_m = \frac{1}{\frac{N}{2} + 1} \sum_{k=1}^{N/2} \log E[\omega_k] \cos(m\omega_k), \quad (1.4)$$

где: c_m - кепстральные коэффициенты с индексами m .

Далее вычисляется искомая частотная характеристика фильтра по формуле:

$$H_s[\omega, n] = \left(\frac{E[\omega, n]}{T[\omega, n]} \right)^{\beta P_v[n]} \quad (1.5)$$

Значение параметра β принимается равным 0.25 для малых соотношений сигнал/ шум, однако оно может быть уменьшено для менее интенсивных шумов $v[n]$. Частотная характеристика второго адаптивного фильтра $H_p[\omega_k]$ вычисляется следующим образом:

$$H_p[\omega_k, n] = \begin{cases} 1, & \omega_k < \omega_0 \\ 1 + \frac{\omega_k - \omega_0}{\pi - \omega_0} g P_v[n], & \omega_k \geq \omega_0 \end{cases} \quad (1.6)$$

где: $\omega_0 = 0,125\pi$ при частоте дискретизации 16кГц.

Значение коэффициента g может быть постоянным (например, 0,3) либо может изменяться в зависимости от соотношения сигнал/шум. Частотная характеристика фильтра с постоянными параметрами $H_p[\omega_k]$ рассчитывается таким образом, чтобы усилить энергию сигнала в диапазоне от 1 кГц до 4 кГц на 12 дБ и ослабить частоты ниже 500 Гц на 6 дБ/октава.

Измененный амплитудный спектр объединяется с исходным фазовым спектром и вычисляется обратное преобразование Фурье. После этого фреймы полученного сигнала суммируются с перекрытием и формируется обработанный

сигнал - полезный сигнал $x[n]$ с измененной формой спектральной огибающей, подчеркивающей формантную структуру речевого сигнала.

3. Выполняется расширение и компрессия динамического диапазона сигнала $x[n]$, полученного на предыдущем шаге. Для этого вычисляется его амплитудная огибающая $\hat{e}[n]$ при помощи преобразования Гильберта. Затем выполняется компрессия с коротким временем отпускания (примерно 2мс) и почти мгновенным временем срабатывания:

$$\hat{e}[n] = \begin{cases} a_r \hat{e}[n-1] + (1-a_r)e[n], & e[n] < \hat{e}[n-1] \\ a_\alpha \hat{e}[n-1] + (1-a_\alpha)e[n], & e[n] \geq \hat{e}[n-1] \end{cases} \quad (1.7)$$

$$a_r=0,15 \text{ и } a_\alpha=0,0001.$$

Затем сглаженная амплитудная огибающая $\hat{e}[n]$ переводится в децибелы $e_{in}[n]=20\log_{10}(\hat{e}[n]/e_0)$, где индекс in обозначает огибающую входного сигнала компрессора, и при помощи кривой амплитудной характеристики компрессора (рисунок 8) вычисляются значения огибающей $e_{out}[n]$ на его выходе. Величина e_0 является опорной величиной, соответствующей 0 дБ.

Амплитудная характеристика компрессора (зависимость уровня амплитудной огибающей $e_{out}[n]$ на выходе компрессора от уровня на входе $e_{in}[n]$) показывает, что при низких значениях амплитуды входного сигнала $x[n]$ (менее -30 дБ) усиления не происходит - выходной уровень равен входному. При умеренных значениях амплитудной огибающей входного сигнала (от -30 до 0 дБ) происходит его усиление - соответствующая область кривой является областью расширения. При высоких значениях амплитудной огибающей входного сигнала $x[n]$ (более 0 дБ) выполняется ослабление - область компрессии. Коэффициент усиления сигнала $x[n]$ вычисляется следующим образом:

$$g[n]=10^{(e_{out}[n]-e_{in}[n])/20} \quad (1.8)$$

При этом выходной сигнал $y[n]$ блока обработки (рисунок 6) формируется путем умножения обработанного сигнала $x[n]$ с измененной формой спектральной огибающей, подчеркивающей формантную структуру

речевого сигнала, на полученный коэффициент усиления $y[n]=g[n]x[n]$ и нормирования его для сохранения исходной энергии.

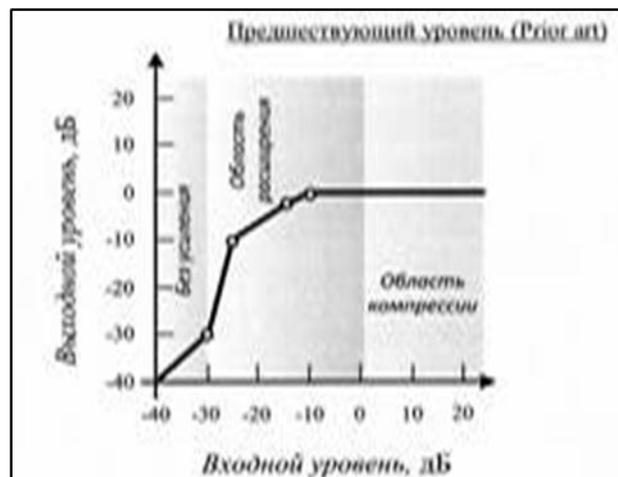


Рисунок 8. Амплитудная характеристика компрессора, предшествующий уровень

Анализ данного способа повышения разборчивости речевых сигналов - полезных сигналов $x[n]$ в шумовой обстановке показывает, что способ предназначен для обработки исключительно речевых сигналов и предполагает наличие в сигнале специфических для речи характеристик, таких как формантные частоты и вокализованные фрагменты, а также подразумевает возможность произвести классификацию диктора (мужчина либо женщина). Способ не применим к обработке неречевых звуковых сигналов (например, музыкальных произведений). Способ сохраняет исходную энергию речевого сигнала, вследствие чего невозможно обеспечить удовлетворительную разборчивость при низких соотношениях сигнал/шум. Даже при наличии идеальных условий (в полной тишине) в обрабатываемый сигнал вносятся искажения. Способ лишь ограниченно учитывает интенсивность шума $v[n]$ и совсем не учитывает его кратковременные спектральные и динамические характеристики. В результате способ плохо применим для нестационарных шумов $v[n]$.

Известен способ повышения разборчивости речевого сигнала - полезного сигнала $x[n]$ путем усиления спектральных составляющих речевого сигнала

таким образом, чтобы для каждой из них обеспечивалось требуемое соотношение сигнал/шум [59]. В общем, алгоритм обработки в этом способе сводится к следующим шагам:

1. В блоке обработки (рисунок 3) вычисляются спектры полезного $x[n]$ и шумового сигналов $v[n]$ при помощи кратковременного преобразования Фурье.

2. Вычисляются соотношения сигнал/шум для каждой спектральной составляющей и требуемые коэффициенты усиления (коэффициенты вычисляются независимо друг от друга).

3. Спектральные составляющие полезного сигнала $x[n]$ умножаются на соответствующие им коэффициенты и переводятся во временную область при помощи обратного преобразования Фурье для формирования выходного сигнала $y[n]$ блока обработки.

Недостатком этого способа является то, что в результате обработки в полезном сигнале $x[n]$ происходит выравнивание амплитуды тихих звуков с потерей их относительной громкости, что приводит к неестественному звучанию. Усиление спектральных компонент выполняется независимо друг от друга, что приводит к потере естественной огибающей обрабатываемого сигнала. Еще одним ограничением способа является то, что он не предназначен для обработки музыкальных произведений, поскольку целью является повышение разборчивости речевого сигнала, имеющего менее широкий динамический диапазон, чем музыкальный диапазон.

1.4. ПОСТАНОВКА ЗАДАЧИ ИССЛЕДОВАНИЯ

Проведенные анализ способов повышения разборчивости речи для слабослышащих абонентов в цифровых телекоммуникационных сетях и анализ способов уменьшения влияния шума окружающей среды на разборчивость звуковых сигналов при использовании мобильных устройств телекоммуникации показали, что проблема клиенто-ориентированного доступа слабослышащих

абонентов к ресурсам сетей, систем и устройств телекоммуникаций недостаточно решена. Имеющиеся решения ориентированы на прием слабослышащими абонентами только речевой информации и не предназначены для обработки приема слабослышащими абонентами более сложных звуковых композиций, например, музыкальных произведений.

В результате анализа установлено, что известные слуховые аппараты мало приспособлены к индивидуальным особенностям слуха слабослышащего абонента и не ориентированы на обеспечение доступа к ресурсам сетей, систем и устройств телекоммуникаций. Поэтому для повышения разборчивости и распознавания звуковых сигналов и доступа слабослышащих абонентов к ресурсам сетей, систем и устройств телекоммуникаций необходимы новые способы повышения разборчивости, информативности и натуральности звучания звуковых сигналов, в том числе и музыкальных композиций при прослушивании в акустической шумовой обстановке, ориентированные на индивидуального абонента и особенности его слуха.

Современные информационные технологии, в первую очередь, методы искусственного интеллекта и машинного обучения, такие как методы автоматизированного распознавания и преобразования устной речи в текст также могли бы быть применены для повышения разборчивости и распознавания звуковых сигналов и доступа слабослышащих абонентов к ресурсам сетей, систем и устройств телекоммуникаций. Для этого необходимо проанализировать возможность использования систем автоматизированного распознавания речи в мобильных устройствах сетей телекоммуникации и решить проблему уменьшения информационных и вычислительных ресурсов.

На основании проведенного анализа цель исследования можно сформулировать как разработку комплексной технологии и устройств, обеспечивающих клиенто-ориентированный доступ слабослышащих абонентов к ресурсам сетей, систем и устройств телекоммуникаций.

Для достижения поставленной цели потребуется разработать новые пути повышения разборчивости звуковых сигналов и доступа слабослышащих абонентов к ресурсам сетей, систем и устройств телекоммуникаций при прослушивании в акустической шумовой обстановке и решить ряд задач, основными из которых являются:

- разработка метода и программно-аппаратных средств формирования персонализированных аудиосигналов для слабослышащих пользователей на основе их частотных характеристик слуха;
- разработка метода уменьшения влияния шума окружающей среды для прослушивания звуковых сигналов разного типа: устной речи, музыкальных композиций;
- разработка программно-аппаратного комплекса для совмещения функционала мобильного и слухового аппарата на базе разработанных методов;
- разработка метода формирования базы данных сбалансированных синтагм для автоматизированного распознавания, синтеза и преобразования устной речи в текст;
- создание аппаратно-программного обеспечения устройств автоматизированной транскрипции устной речи, снижающее необходимые информационные и вычислительные ресурсы при использовании слабослышащими абонентами мобильных телефонов.

Выводы по главе 1.

1. Потерей слуха страдает 1,5 миллиарда человек в мире, однако в настоящее время производство слуховых аппаратов удовлетворяет менее 10% мировых потребностей.

2. Обеспечение равного доступа людям с ограниченными возможностями к ресурсам сетей, информационных систем и устройств телекоммуникации является важной экономической, социальной и гуманитарной задачей любого

современного общества и является основой для развития ассистивных технологий.

3. Существующие устройства компенсации слуха (слуховые аппараты – СА) не только являются дефицитными и дорогостоящими, но и не отвечают по своим техническим характеристикам требованиям разборчивости речи в условиях агрессивного шумового окружения.

4. Существующие методы шумоочистки речи предназначены, как правило, для обработки исключительно речевых сигналов и предполагают наличие в сигнале специфических для речи характеристик, но не позволяют обрабатывать неречевые звуковые сигналы (например, музыкальные произведения), не обеспечивают удовлетворительную разборчивость при низких соотношениях сигнал/шум; естественность передачи речи; плохо применимы для нестационарных шумов.

5. Существующие способы повышения разборчивости речи у пользователей с нейросенсорной тугоухостью в цифровых телекоммуникационных системах, базирующийся на слуховом аппарате (СА) и мобильном телефонном аппарате (ТА) не решают проблемы как акустической совместимости, так и электромагнитной совместимости данных двух приборов.

6. Современными смартфонами могут стать эффективным решением для диагностики и компенсации нарушений слуха при разработке способа шумоочистки, обеспечивающего удовлетворительную разборчивость при низких соотношениях сигнал/шум; естественность передачи речи; применимость для нестационарных шумов, а также способа акустической и электромагнитной совместимости слухового аппарата и мобильного телефона или альтернативного ему.

7. Для оснащения мобильных телефонов с целью компенсации потери слуха необходимо разработать специальное аппаратное и программное обеспечение, решающее проблемы:

- невозможности использования пользователем устройства персонализированного аудиосигнала, необходимого во время личного общения с человеком-собеседником;

- отсутствия возможности пользователем устройства получить персонализированный аудиосигнал при одновременном общении пользователя с другим абонентом по телефону и с собеседником лично, при приеме аудиосигнала из динамиков различных устройств, например во время просмотра передач телевидения, прослушивания музыки и т.п.;

- неосуществимость получения персонализированного аудиосигнала слабослышащими абонентами сетей связи, не имеющими подобных устройств, непосредственно из сетей связи, в частности при общении пользователя известного устройства с другим слабослышащим абонентом сети связи, не обладающего подобным устройством;

- отсутствие у пользователя устройства в соответствии с его предпочтениями реализации различных режимов: телефонного разговора, личного общения с собеседником, приема аудиосигнала из динамиков разнообразных устройств.

Таким образом, выполненный анализ показал необходимость в разработке новых методов повышения разборчивости речи в условиях агрессивной шумовой обстановке и совмещения функциональных возможностей слухового и мобильного аппаратов для реализации устройств, обеспечивающих слабослышащим клиентам доступ к ресурсам сетей, систем и устройств телекоммуникаций. Рассмотрим разработку таких методов и устройств в следующей главе.

ГЛАВА 2. РАЗРАБОТКА СПОСОБОВ И УСТРОЙСТВ, ОБЕСПЕЧИВАЮЩИХ КЛИЕНТО-ОРИЕНТИРОВАННЫЙ ДОСТУП К РЕСУРСАМ СЕТЕЙ, СИСТЕМ И УСТРОЙСТВ ТЕЛЕКОММУНИКАЦИЙ

2.1 РАЗРАБОТКА И ОПИСАНИЕ СПОСОБА ПОВЫШЕНИЯ РАЗБОРЧИВОСТИ И ПОНЯТНОСТИ ЗВУКОВЫХ СИГНАЛОВ РАЗНОГО ТИПА СЛОЖНОСТИ В ШУМОВОЙ ОБСТАНОВКЕ

Разработанный и представленный в работе способ повышения разборчивости речи опирается на следующее техническое решение [59]. Входной речевой сигнал блока обработки (рисунок 6) – «полезный акустический сигнал» $x[n]$. Речевой $x[n]$ и шумовой $v[n]$ сигнал делятся на фреймы длиной N отсчетов с перекрытием в половину длины фрейма (один фрейм соответствует 20 мс): $x^{(i)}[n], v^{(i)}[n], 1 \leq n \leq N$, где i - индекс фрейма обработки.

Каждый фрейм умножается на окно Хэннинга и переводится в частотную область при помощи дискретного преобразования Фурье. В результате формируются кратковременные спектры речевого и шумового сигналов, обозначенные далее $x^{(i)}[\Omega_\mu]$ и $v^{(i)}[\Omega_\mu]$ соответственно, где Ω_μ - дискретная частота, а μ - индекс частоты. Спектр речевого сигнала $x[n]$ умножается на коэффициенты усиления $G^{(i)}[\Omega_\mu]$ и вычисляется измененная амплитуда $\hat{x}^{(i)}[\Omega_\mu]$:

$$\hat{x}^{(i)}[\Omega_\mu] = G^{(i)}[\Omega_\mu] x^{(i)}[\Omega_\mu]. \quad (2.1)$$

Коэффициенты усиления $G^{(i)}[\Omega_\mu]$ принимают значения большие или равные единице и вычисляются таким образом, чтобы обеспечить соотношение сигнал/шум для каждого частотного отсчета не ниже заданной величины. Таким образом, выполняется ограничение минимального выходного уровня сигнала $y[n]$ на выходе блока обработки в зависимости от уровня шума $v[n]$. После умножения на коэффициенты усиления полученный образ Фурье с измененной амплитудой $\hat{x}^{(i)}[\Omega_\mu]$ переводится во временную область при помощи обратного

преобразования Фурье. Из полученных фреймов формируется выходной сигнал блока обработки $y[n]$ методом сложения с перекрытием.

Компрессор-ограничитель уровня выходного сигнала блока обработки $y[n]$ имеет два входа - спектр полезного сигнала $x[n]\hat{x}^{(i)}[\Omega_\mu]$ и спектр $v^{(i)}[\Omega_\mu]$ сигнала шума $v[n]$. На выходе блока обработки имеем сигнал $y[n]$ с измененной амплитудой $\hat{x}^{(i)}[\Omega_\mu]$ - сигнал на входе громкоговорителя, который воспринимает «Слушатель» (рисунок 6). Обработка сигналов полезного $x[n]$ и шума $v[n]$ в блоке обработки ведется последовательно - фрейм за фреймом по следующим шагам:

1. Вычисляются средние значения спектральной плотности мощности речевого(полезного) сигнала $x[n]$ и шума $v[n]$, обозначенные соответственно $x^{(i)}_{nos}[\Omega_\mu]$ и $x^{(i)}_{sig}[\Omega_\mu]$:

$$x^{(i)}_{sig}[\Omega_\mu] = \alpha_s x^{(i-1)}_{sig}[\Omega_\mu] + (1 - \alpha_s) x^{(i-1)}[\Omega_\mu]^2 \quad (2.2)$$

$$x^{(i)}_{nos}[\Omega_\mu] = \alpha_N x^{(i-1)}_{nos}[\Omega_\mu] + (1 - \alpha_N) v^{(i)}[\Omega_\mu]^2, \quad (2.3)$$

где $\alpha_s, \alpha_N \in [0, 1]$ являются коэффициентами экспоненциального усреднения и рекомендуется использовать следующие значения

$$\alpha_s = 0,996 \text{ и } \alpha_N = 0,96.$$

2. Вычисляются коэффициенты усиления

$$G^{(i)}[\Omega_\mu] = \min \left\{ \max \left\{ \sqrt{\xi \frac{x^{(i)}_{nos}[\Omega_\mu]}{x^{(i)}_{sig}[\Omega_\mu]}}, 1 \right\}, G_{\max} \right\}, \quad (2.4)$$

где ξ - заданное минимальное соотношение сигнал/шум;

G_{\max} - ограничение максимального усиления.

3. Ограничение максимальной амплитуды спектральных составляющих выходного сигнала $y[n]$ для предотвращения превышения болевого порога

$$\hat{x}^{(i)}[\Omega_\mu] = \begin{cases} \hat{x}_{\max}[\Omega_\mu] \frac{\hat{x}^{(i)}[\Omega_\mu]}{|x^{(i)}[\Omega_\mu]|}, & \text{если } |x^{(i)}[\Omega_\mu]| > \hat{x}_{\max}[\Omega_\mu], \\ \hat{x}^{(i)}[\Omega_\mu], & \text{если } |x^{(i)}[\Omega_\mu]| \leq \hat{x}_{\max}[\Omega_\mu] \end{cases}, \quad (2.5)$$

где $\hat{x}_{\max}[\Omega\mu]$ - максимально допустимые значения амплитуды спектральных составляющих.

Недостатком этого решения является то, что он позволяет повысить разборчивость речи - полезного сигнала $x[n]$, но без учета комфорта восприятия, естественности и натуральности звучания выходного сигнала $y[n]$ на фоне шумов. В результате данное решение применимо в каналах речевой связи, однако неприменимо для прослушивания при внешнем шуме музыкальных произведений, таких как музыка с речевым сопровождением, например песни, или без речевого сопровождения, например пьесы, симфонии и т.п. Предполагается, что требуемая разборчивость получается при достижении некоторого заданного минимального соотношения сигнал/шум для всех звуков слышимого динамического диапазона. Для чего используется компрессор-ограничитель минимального выходного уровня сигнала $y[n]$ (рисунок 9).

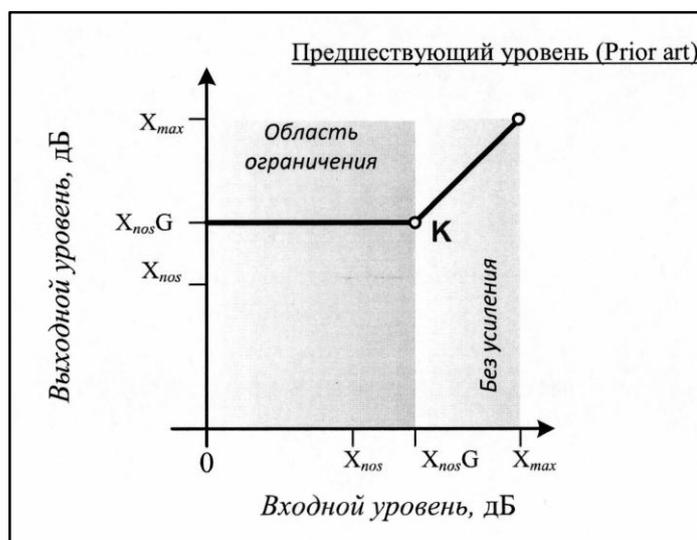


Рисунок 9. Амплитудная характеристика компрессора-ограничителя уровня выходного сигнала $y[n]$, предшествующий уровень

Использование ограничителя минимального выходного уровня сигнала $y[n]$ с постоянным выходным уровнем в области ограничения (рисунок 9, точка K на амплитудной характеристике компрессора-ограничителя является точкой перегиба характеристики) приводит к потере натуральности звучания, поскольку в области ограничения теряется градация тихих звуков - на выходе все звуки

становятся одной интенсивности. В то же время для сохранения динамических оттенков и повышения натуральности звучания необходимо использовать изменяющийся (от тихого к громкому) выходной уровень в области ограничения.

Можно отметить, что в этом техническом решении усиление каждого частотного отсчета спектра сигнала $x[n]$ выполняется независимо друг от друга. Это приводит к потере относительной громкости между частотными составляющими сигнала, теряется форма его спектральной огибающей и в результате сильно искажается тембр, как показано на рисунке 10.

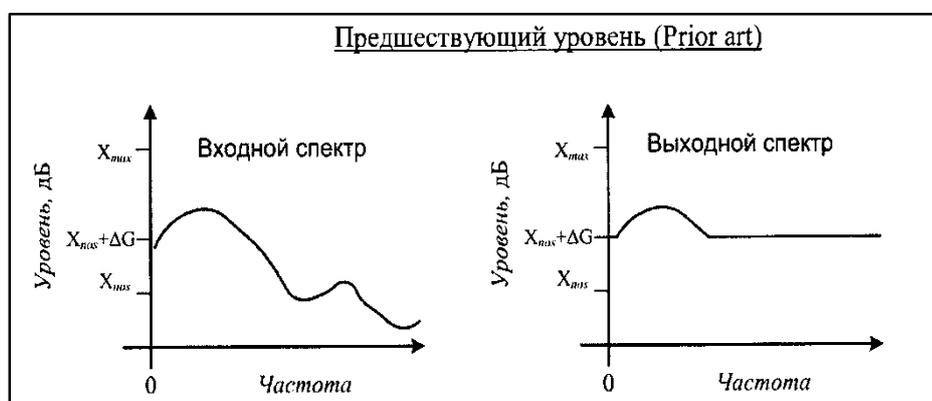


Рисунок 10. Входной и выходной спектры полезного акустического сигнала $x[n]$ для случая белого шума $v[n]$, предшествующий уровень

Спектральный анализ полезного сигнала $x[n]$ и шума $v[n]$ в данном способе выполняется с равным частотным разрешением по всей шкале частот (обусловлено применением дискретного преобразования Фурье) без учета особенностей частотного разрешения слуховой системы слушателя.

Таким образом, этот известный способ повышения разборчивости и информативности звуковых сигналов в шумовой обстановке заключается в том, что:

- полезный сигнал $x[n]$, поступающий в блок обработки, и шумовой $v[n]$ сигнал акустической обстановки, поступающий из окружающего пространства в блок обработки, делят на фреймы длиной по N отсчетов с

перекрытием в половину длины фрейма, где N - размер фрейма, а n - номер отсчета сигнала во фрейме, $n=1,2\dots N$;

- обработку фреймов производят в блоке обработки фрейм за фреймом, каждый фрейм умножаем на окно Хэннинга и производим декомпозицию полезного сигнала $x[n]$ и шумового сигнала $v[n]$ акустической обстановки на частоты и амплитуды путем дискретного преобразования Фурье, в результате чего формируются кратковременные спектральные зависимости полезного сигнала и шумового сигнала акустической обстановки,

- изменяют динамический диапазон полезного сигнала и шумового сигнала акустической обстановки в компрессоре блока обработки, для чего кратковременные спектральные зависимости полезного сигнала $x[n]$ умножают на коэффициенты усиления $g[n] \geq 1$, обеспечивающие заданное соотношение сигнал/шум для каждого частотного отсчета n ,

- после изменения динамического диапазона производят композицию измененных кратковременных спектральных зависимостей полезного сигнала путем обратного преобразования Фурье, получая скорректированные фреймы, из которых формируют путем сложения с перекрытием выходной сигнал $y[n]$, который из блока обработки передают в окружающее пространство.

Решаемая разработанным способом задача - повышение технико-эксплуатационных характеристик, расширение функциональных возможностей мобильных мультимедийных устройств (смартфонов и планшетов), мультимедийных воспроизводящих устройств, систем видеоконференций, громкой связи, IP - телефонии, колл-центров, усовершенствование слуховых аппаратов, корректирующих звуковой сигнал (во время телефонного разговора и при прослушивании аудиоконтента) в агрессивной акустической шумовой обстановке.

Для решения этой задачи в описанном способе повышения разборчивости и информативности звуковых сигналов в шумовой обстановке:

- в блок обработки вводят банки фильтров анализа и банки фильтров синтеза, которыми производят субполосную декомпозицию полезного сигнала и шума акустической обстановки и, соответственно, субполосную композицию субполос полезного сигнала,
- при субполосной декомпозиции вычисляют энергию в каждой субполосе полезного сигнала $x[n]$ и сигнала шума $v[n]$ акустической обстановки для расчета коэффициентов усиления в каждой субполосе,
- в качестве компрессора блока обработки используют адаптивный компрессор динамического диапазона, которым изменяют динамический диапазон полезного сигнала, для чего сигналы в субполосах полезного сигнала умножают на коэффициенты усиления в соответствии с амплитудной характеристикой адаптивного компрессора динамического диапазона в каждой субполосе,
- при этом положение точки перегиба на амплитудной характеристике адаптивного компрессора динамического диапазона, разделяющей динамический диапазон полезного сигнала на область компрессии и на область - без усиления, обуславливают уровнем энергии субполосного сигнала шума акустической обстановки,
- в соответствии с амплитудной характеристикой области компрессии адаптивного компрессора динамического диапазона в каждой субполосе полезного сигнала определяют коэффициент усиления на основе уровня энергии субполосных сигналов полезного сигнала,
- после субполосной композиции банками фильтров синтеза измененных амплитудных зависимостей полезного сигнала из скорректированных фреймов формируют путем сложения со стыковкой выходной сигнал $y[n]$.

Возможны дополнительные варианты осуществления способа, в которых целесообразно, чтобы:

- в качестве банка фильтров использовали M -канальный неравнополосный косинусно-модулированный банк фильтров;

- для устранения артефактов в выходном сигнале в блоке обработки для каждого фрейма полезного сигнала выполняли интерполяцию коэффициентов усиления адаптивного компрессора динамического диапазона с задержкой на размер фрейма обработки, причем состыкованные для соседних фреймов коэффициенты усиления представляют в виде кусочно-линейной функции, получая плавную регулировку уровня громкости выходного сигнала блока обработки.

Таким образом, существенными отличиями разработанного способа является то, что:

- декомпозицию полезного сигнала $x[n]$ и шумового $v[n]$ сигнала производят не путем дискретного преобразования Фурье, а композицию (синтез) не обратным преобразованием Фурье, а банками фильтров анализа и банками фильтров синтеза, которыми производят субполосную обработку,

- используют не компрессор-ограничитель уровня выходного сигнала $y[n]$, адаптивный компрессор динамического диапазона, в котором положение точки K перегиба на его амплитудной характеристике компрессора, разделяющей динамический диапазон входного сигнала $x[n]$ на область компрессии и на область - без усиления, обуславливают уровнем энергии шумового сигнала $v[n]$;

- выходной сигнал $y[n]$ формируют не путем сложения с перекрытием из скорректированных фреймов, а их сложением со стыковкой.

Достоинством предложенного способа повышения разборчивости и информативности звуковых сигналов в шумовой обстановке является улучшение восприятия различных звуковых (в том числе речевых) сигналов на фоне акустических шумов $v[n]$, причем сохранение натуральности и естественности звучания является основным преимуществом. При обработке усиливаются лишь

те фрагменты полезного сигнала $x[n]$, которые маскируются шумом $v[n]$. В полной тишине в полезный сигнал $x[n]$ не вносятся никаких изменений.

Динамический диапазон полезного сигнала $x[n]$ сужается при помощи адаптивного (к шуму) компрессора. В результате обеспечивается хорошая слышимость в полезном сигнале $x[n]$ как изначально тихих, так и громких звуков без превышения допустимой громкости и с сохранением относительных градаций от тихого к громкому.

Субполосная декомпозиция полезного сигнала $x[n]$ и шума $v[n]$ выполняется на основе банка фильтров (неравнополосного), адаптированного к неравномерной частотной шкале слухового восприятия. Обработка во временной области, а не в частотной, обеспечивает сохранение оригинальной формы спектральной огибающей сигнала в каждой из частотных полос банка фильтров.

Энергия выходного сигнала $y[n]$ непосредственно зависит от энергии шума $v[n]$, что позволяет обеспечить хорошую слышимость звука даже при очень низких соотношениях сигнал/шум. Усиление полезного сигнала $x[n]$ выполняется в соответствии с кратковременными спектральными и динамическими характеристиками шума $v[n]$, что делает изложенный способ пригодным для использования в условиях нестационарных шумов $v[n]$.

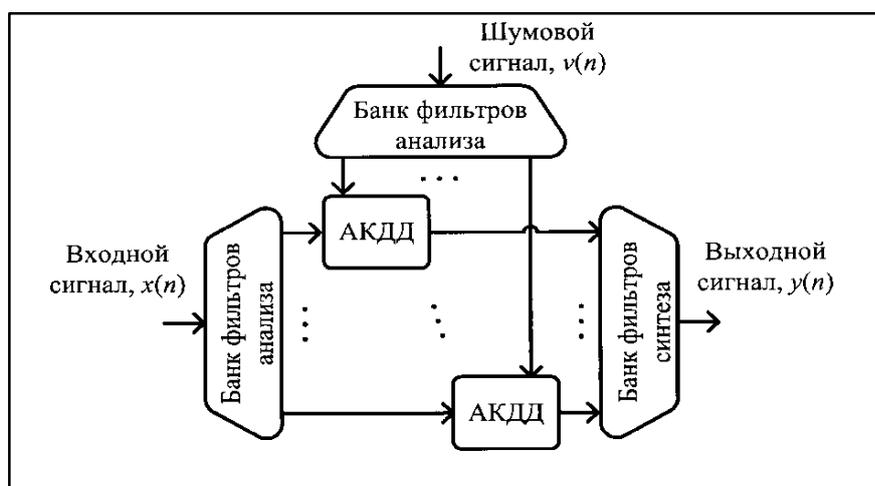


Рисунок 11. Субполосная декомпозиция полезного сигнала $x[n]$, шума $v[n]$ и формирование выходного сигнала $y[n]$ в блоке обработки

Для реализации разработанного способа в блоке обработки (рисунок 6) выполняется субполосная декомпозиция (рисунок 11) полезного сигнала $x[n]$ и сигнала шума $v[n]$, принятого из микрофона. Для обработки субканальных сигналов используются адаптивные компрессоры динамического диапазона (АКДД).

Чтобы учесть спектральные особенности шума $v[n]$ окружающей обстановки и выполнить усиление только тех частотных областей полезного сигнала $x[n]$, которые маскируются шумом, в предлагаемом способе может использоваться неравнополосный косинусно-модулированный банк фильтров.

Банк фильтров позволяет выполнить субполосную декомпозицию сигналов схожую с той, которая выполняется улиткой уха человека, и осуществлять обработку сигналов во временной области, что позволяет избежать ошибок при переходе из частотной области во временную. Банки фильтров анализа для полезного сигнала $x[n]$ и шумового сигнала $v[n]$ выполнены одинаково. В каждом канале выполняется корректировка громкости для повышения уровня разборчивости и информативности полезного сигнала $x[n]$ с учетом шума $v[n]$ окружающей среды. Для обработки субканальных сигналов используется адаптивный компрессор динамического диапазона. В каждом канале обработка ведется независимо. Формирование выходного сигнала $y[n]$ блока обработки (рисунок 6) выполняется с помощью банка фильтров синтеза (рисунок 11). Процесс обработки сигналов полезного $x[n]$ и шума $v[n]$ в блоке обработки (рисунки 6 и 11) ведется фрейм за фреймом.

Адаптивный компрессор динамического диапазона (АКДД) имеет два входа - для полезного сигнала $x[n]$, и сигнала шума $v[n]$, и один выход $y[n]$ - сигнал на входе громкоговорителя, который воспринимает «Слушатель». Целью АКДД является расчет таких коэффициентов усиления $g[n]$, которые не позволят шуму маскировать полезный сигнал во всем частотном диапазоне изменения шумового сигнала $v[n]$:

$$y[n]=x[n]\cdot g[n] \quad (2.6)$$

Одной из особенностей работы АКДД является то, что для избежания появления артефактов в выходном сигнале $y[n]$ коэффициенты усиления $g[n]$, состыкованные для соседних фреймов, должны представлять собой кусочно-линейную функцию без разрывов. Исходя из этого требования, создается специальная характеристика адаптивного к шуму $v[n]$ компрессора динамического диапазона, работа которого складывается из следующих шагов:

1. Канальные сигналы полезного $x^{(i)}[n]$ и шума $v^{(i)}[n]$, $n=0, 1 \dots N$, где i - индекс фрейма, а N - размер фрейма (число отсчетов сигнала во фрейме), поступают на входы соответствующих АКДД (рисунок 11).

2. В АКДД для шумового сигнала $v^{(i)}[n]$ сохраняется значение уровня энергии шума для предыдущего кадра, новое значение уровня энергии шума рассчитывается следующим образом:

$$V^{(i)}=10\log_{10}\left(\frac{1}{N}\sum_{n=0}^{N-1}(v^{(i)}[n])^2\right), \quad (2.7)$$

$$x_{nos}^{(i)} = \varepsilon x_{nos}^{(i-1)} + (1 - \varepsilon)v^{(i)}, \quad (2.8)$$

где $0<\varepsilon<1$ - коэффициент экспоненциального усреднения, величина которого определяет скорость сходимости процедуры усреднения.

3. Значение $x_{nos}^{(i)}$ используется для корректировки амплитудной характеристики АКДД (рисунок 11). Положение точки перегиба K амплитудной характеристики АКДД) меняется в зависимости от текущего уровня шума x_{nos} для каждого фрейма. Если уровень шума x_{nos} увеличивается, то точка K смещается вверх по диагонали (обозначенной пунктирной линией на рисунке 12 самым мелким штрих пунктиром).

Если уровень энергии входного сигнала выше $x_{nos}+\Delta G$, то в АКДД не происходит усиления сигнала, в противном случае уровень энергии выходного сигнала блока обработки рассчитывается, используя нижнее колено амплитудной характеристики АКДД (расположенное в области компрессии).

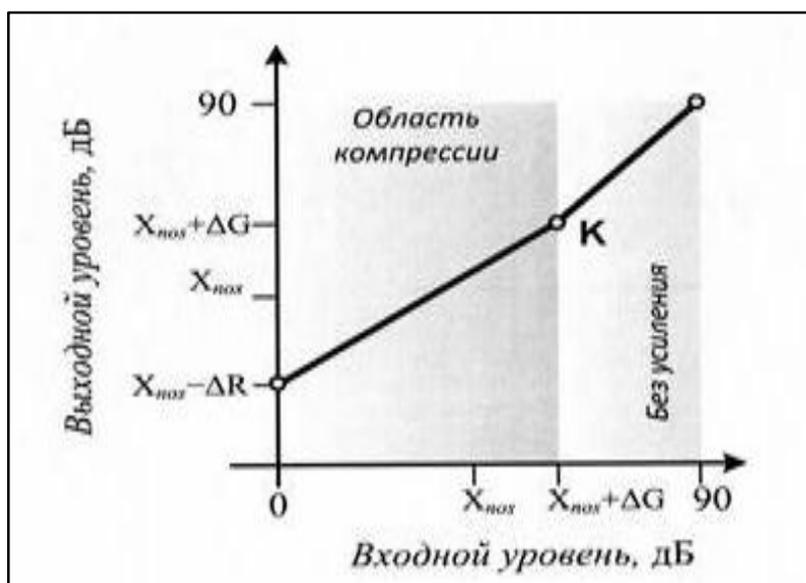


Рисунок 12. Амплитудная характеристика адаптивного компрессора динамического диапазона

Константы ΔG и ΔR являются настроечными параметрами и характеризуют степень вносимого эффекта компрессии. Параметры подбираются исходя из характеристик микрофона, используемого для регистрации шума и, например, если полезный сигнал $x[n]$ - сигнал проигрывателя, качества записи данного сигнала.

4. Определяется для i -го фрейма полезного сигнала $x[n]$ коэффициент усиления (целевой). При этом в АКДД рассчитывается уровень энергии фрейма полезного сигнала $x[n]$.

$$x^{(i)} = 10 \log_{10} \left(\frac{1}{N} \sum_{n=0}^{N-1} (x^{(i)}[n])^2 \right) \quad (2.9)$$

Это значение используется для расчета уровня выходного сигнала $y[n]$:

$$x_{new}^{(i)} = \Psi \left(x^{(i)}, x_{nos}^{(i)} \right), \quad (2.10)$$

где $\Psi \left(x^{(i)}, x_{nos}^{(i)} \right)$ - амплитудная характеристика АКДД (рисунок 12).

Коэффициент усиления определяется для i -го фрейма следующим образом:

$$g^{(i)} = 10^{(x_{new}^{(i)} - x^{(i)})/20} \quad (2.11)$$

5. Выполняется интерполяция коэффициентов усиления $g[n]$ для получения "плавной" регулировки уровня громкости выходного сигнала $y[n]$. Коэффициент $g^{(i)}$ связывается с серединой фрейма, поэтому для расчета всех коэффициентов $g^{(i)}[n]$ необходимо знать значение коэффициента усиления для следующего фрейма $g^{(i+1)}$. Это приводит к тому, что в АКДД вносится задержка на размер фрейма. Для того, чтобы рассчитать усиление текущего фрейма, нужно дождаться следующего фрейма, поэтому обработать и вывести текущий фрейм можно только с опозданием на один на фрейм сигнала. На рисунке 13 поясняется принцип интерполяции коэффициентов $g^{(i-1)}$, $g^{(i)}$ и $g^{(i+1)}$ для получения $g^{(i)}[n]$. Применяется линейная интерполяция значений коэффициентов усиления на все отсчеты полезного сигнала $x[n]$, попадающие в диапазон между центрами предыдущего и следующего кадров.

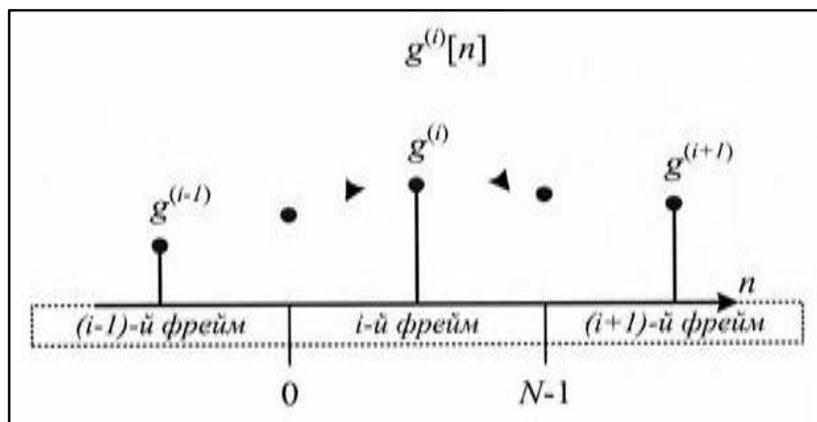


Рисунок 13. Принцип интерполяции коэффициентов $g^{(i-1)}$, $g^{(i)}$ и $g^{(i+1)}$ для получения $g^{(i)}[n]$ для соседних фреймов в адаптивном компрессоре Динамического диапазона.

6. Формирование уровня отсчетов выходного сигнала $y[n]$ блока обработки путем масштабирования уровня отсчетов полезного сигнала $x[n]$ на соответствующий коэффициент $g^{(i)}[n]$ для i -го фрейма полезного сигнала $x[n]$:

$$y^{(i)}[n] = x^{(i)}[n] * g^{(i)}[n], \quad (2.12)$$

где $n=0,1 \dots N-1$

7. Конец работы АКДД.

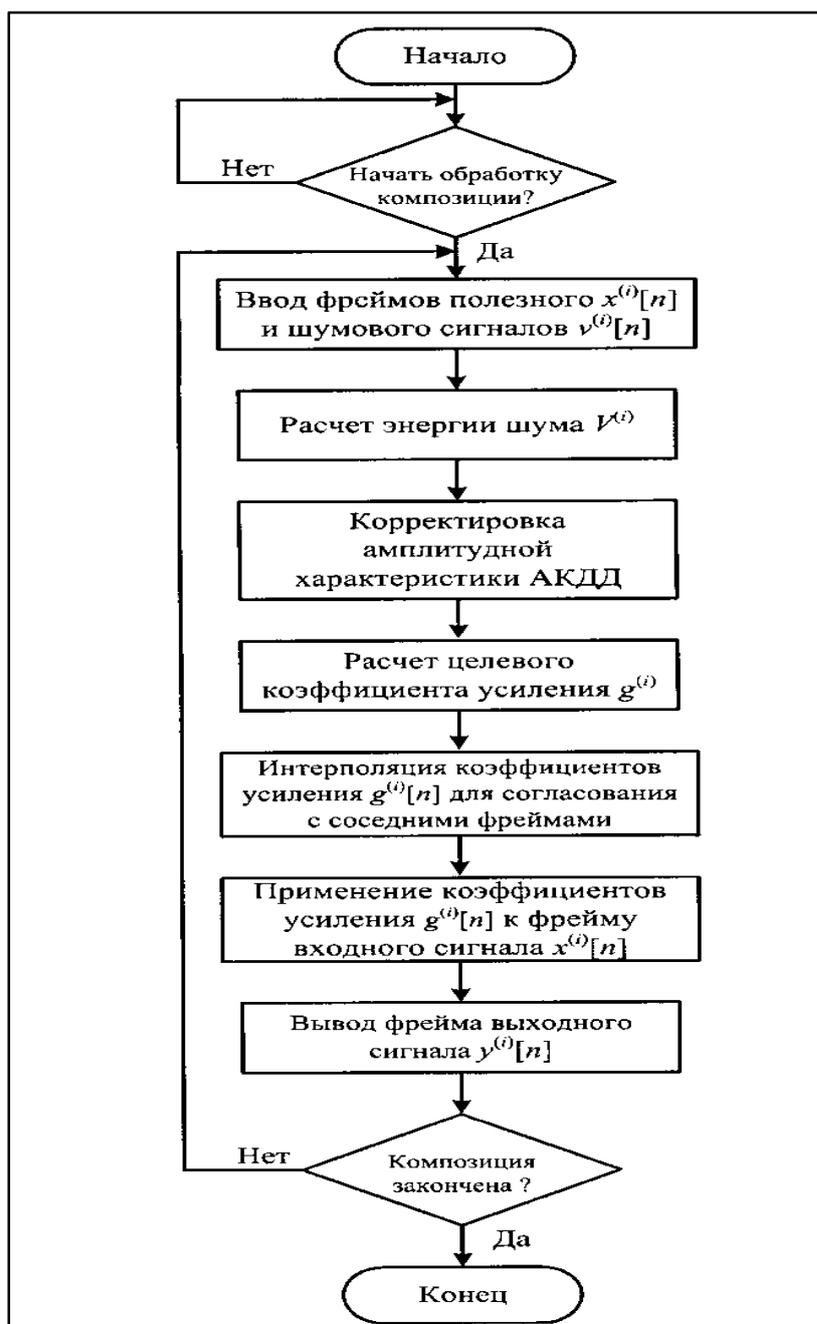


Рисунок 14. Блок-схема алгоритма работы блока обработки

На рисунке 14 представлен алгоритм работы блока обработки. При поступлении полезного сигнала $x[n]$ и шумового сигнала $v[n]$ акустической обстановки в блоке принятия решений «Начать обработку композиции?» производится запуск работы по выходу «Да» программного обеспечения блока обработки.

В результате производится ввод фреймов полезного сигнала $x^{(i)}[n]$ и шумового сигнала $v^{(i)}[n]$ акустической обстановки, где i - индекс фрейма, а n -

номер отсчета. Далее производится расчет энергии шума $V^{(i)}$. В соответствии со значением энергии шума производится корректировка амплитудной характеристики АКДД. Согласно со скорректированной характеристикой АКДД определяется коэффициент усиления $g^{(i)}$. Далее производится интерполяция коэффициентов усиления $g^{(i)}[n]$ для согласования соседних фреймов и устранения артефактов выходного сигнала $y^{(i)}[n]$.

Затем коэффициенты усиления $g^{(i)}[n]$ применяются к фрейму входного полезного сигнала $x^{(i)}[n]$, в результате формируется фрейм выходного сигнала $y^{(i)}[n]$, и осуществляется его вывод. В блоке принятия решения «Композиция закончена?» принимается решение по выходу «Да» на окончание работы, а по выходу «Нет» информация передается на ввод следующих фреймов полезного сигнала $x^{(i)}[n]$ и шумового сигнала $v^{(i)}[n]$ для продолжения обработки.

Субполосная декомпозиция полезного сигнала $x[n]$ и сигнала шума $v[n]$ для расчета маскирующего эффекта шумом $v[n]$ полезного сигнала $x[n]$ может быть эффективно выполнена с использованием 5-канального неравнополосного косинусно-модулированного банка фильтров (НКМБФ), который аппроксимирует шкалу критических частотных полос (психоакустическая частотная шкала Барков) восприятия человеком акустической информации.

Число субполос для реализации предложенного способа определяется качеством формируемого выходного сигнала и вычислительной сложностью реализации банка фильтров. Данный банк фильтров образуется из своего равнополосного аналога путем применения фазового преобразования, т.е. замены всех элементов задержки на фазовые звенья ($z^{-1} \rightarrow A(z)$), где z - оператор Z преобразования. M -канальный НКМБФ задается передаточными функциями канальных фильтров анализа $H_k(z)$ и синтеза $F_k(z)$ следующим образом:

- для банка анализа:

$$H_k(z) = a_k b_k H(A^{-1}(z) W_{2M}^{(k+0,5)}) + \bar{a}_k \bar{b}_k H(A^{-1}(z) W_{2M}^{-(k+0,5)}) \quad (2.13)$$

- для банка синтеза:

$$F_k(z) = \bar{a}_k b_k H\left(A^{-1}(z) W_{2M}^{(k+0,5)}\right) + a_k \bar{b}_k H\left(A^{-1}(z) W_{2M}^{-(k+0,5)}\right), \quad (2.14)$$

где M – число каналов в банке (количество критических частотных полос),

$$a_k = e^{j(-1)^k \pi/4} \quad (2.15)$$

$$b_k = W_{2M}^{\frac{N-1}{2}, (k+0,5)} \quad (2.16)$$

$$W_{2M} = e^{j2\pi/M}, \quad (2.17)$$

где k - номер канала банка,

$j = \sqrt{-1}$ - мнимая единица, а верхняя черта означает комплексно-сопряженное число.

Здесь $H(z)$ - фильтр нижних частот с конечной импульсной характеристикой (КИХ) порядка N и частотой среза $\omega_c = \pi/2M$, на основе которого формируются передаточные функции канальных фильтров анализа $H_k(z)$ и синтеза $F_k(z)$. С целью уменьшения вычислительной сложности НКМБФ используется фазовое звено $A(z)$ первого порядка, передаточная функция которого определяется как

$$A(z) = (\alpha + z^{-1}) / (1 + \alpha z^{-1}), \quad |\alpha| \leq 1 \quad (2.18)$$

$$A(e^{j\omega}) = e^{j\phi(\omega)}, \quad (2.19)$$

где α - коэффициент передаточной функции фазового звена $A(z)$, определяющий его фазочастотную характеристику:

$$\phi(\omega) = -\omega + \arctg(\alpha \sin \omega / (\alpha \cos \omega - 1)). \quad (2.20)$$

Замена ($z^{-1} \rightarrow A(z)$) приводит к деформации оси частот ($\omega \rightarrow \phi(\omega)$) и получению неравнополосного банка фильтров.

Как видно на рисунке 15, степень деформации частотной оси зависит от одного параметра α - коэффициента передаточной функции фазового звена $A(z)$, определяющего фазочастотную характеристику данного звена. Регулируя коэффициент α , можно изменять ширину полос банка фильтров (рисунок 15).

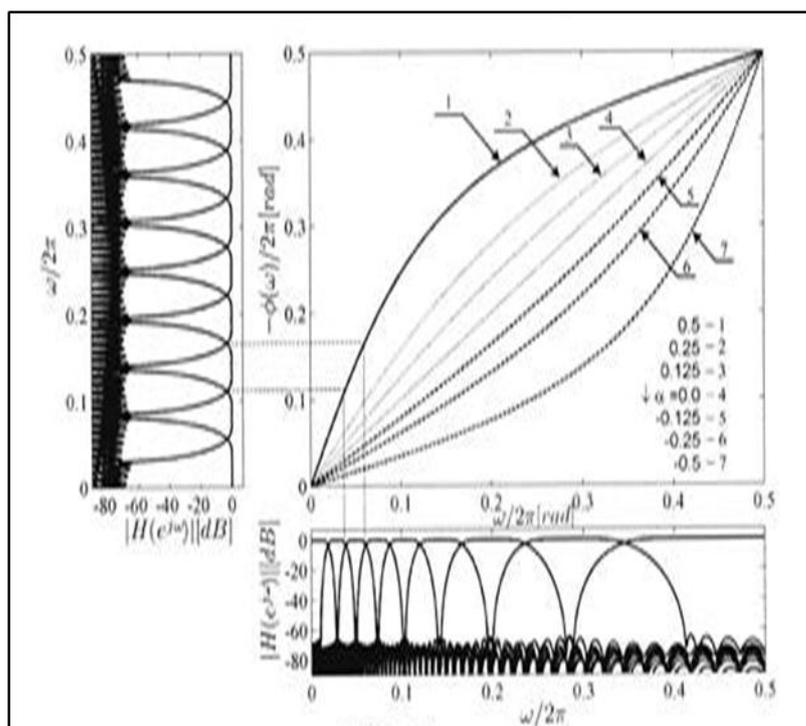


Рисунок 15. Принцип задания неравнополосного банка фильтров на основе деформации частотной оси

В предлагаемом способе используется 5-канальный НКМБФ. Банк фильтров аппроксимирует шкалу Барков; деформация частотной оси задается выбором коэффициента $\alpha=0,7452$, амплитудно-частотная характеристика которого показана на рисунке 16.

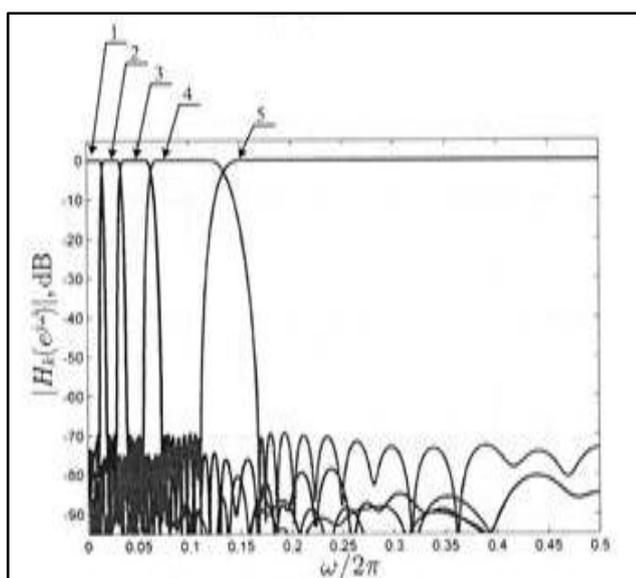


Рисунок 16. Амплитудно-частотная характеристика неравнополосного 5-канального банка фильтров, аппроксимирующая шкалу критических частотных полос (психоакустическую частотную шкалу Барков)

Все пять субполосных фильтров, обозначенные на рисунке 16 как 1, 2, ..., 5 обеспечивают ослабление в полосе не пропуска не менее -70 дБ. При этом погрешность восстановления сигнала банком синтеза не более 0,001 дБ для коэффициентов децимации/интерполяции в каналах банков анализа/синтеза со следующими значениями {21,7,3,1,1}.

2.2 РАЗРАБОТКА СПОСОБА КОМПЕНСАЦИИ ПОТЕРИ СЛУХА С ПОМОЩЬЮ МОБИЛЬНЫХ УСТРОЙСТВ

Разработка предлагаемого решения базировалось на известном способе компенсации потери слуха в телефонной системе, основанном на резолюции номера телефона [58]. В этом способе формируют персонализированный аудиосигнал для слабослышащих пользователей на основе их атрибутов, полученных из их аудиограмм, хранящихся в базе данных и имеющих привязку к телефонным номерам слабослышащих пользователей.

Этот способ реализуется в сети связи, состоящей из ТА ближнего пользователя (абонента) и удаленного пользователя, устройств доступа в сеть данных ТА, а также автоматической телефонной станции - сервера сети, на котором размещается база атрибутов слабослышащих абонентов, программы обработки сигналов ближнего и удаленного абонентов и система селекции атрибутов по номеру слабослышащего пользователя. На сервере связи обрабатывают аудиосигналы в широкополосном частотном диапазоне на основе функции обратной к частотной характеристике слуха слабослышащего пользователя, усиливают и/или ограничивают мощность обработанных аудиосигналов в соответствии с функцией обратной к частотной характеристике слуха слабослышащего пользователя для сохранения умеренной громкости, передают усиленные и/или ограниченные персонализированные аудиосигналы с сервера связи к телефонным аппаратам слабослышащих пользователей. В зависимости от того, кто из абонентов - ближний или удаленный является слабослышащим, можно выделить два варианта реализации способа:

1. ***Удаленный абонент имеет нормальный слух, а ближний абонент - слабослышащий.*** В данном случае способ обработки речевого сигнала следующий. Речевой сигнал ближнего абонента без обработки через устройство доступа в сеть передается по сети связи в устройство доступа в сеть удаленного абонента и далее в ТА удаленного абонента. Аудиосигнал удаленного абонента через устройство доступа в сеть, на основании номера телефона слабослышащего абонента (ближнего пользователя), поступает на сервер сети, где в программном модуле обработки сигнала удаленного абонента подвергается обработке сервером, согласно атрибутам аудиограммы ближнего абонента, которая выбирается из базы атрибутов в соответствии с номером телефона ближнего абонента. Далее обработанный сигнал удаленного абонента через устройство доступа в сеть передается по сети связи в устройство доступа в сеть ближнего абонента и далее в телефонный аппарат ближнего абонента;

2. ***Удаленный и ближний абоненты оба слабослышащие.*** Тогда способ обработки речевых сигналов в коммуникационной сети реализуется следующим образом. Аудиосигналы (речевые) ближнего и удаленного абонентов через соответствующие устройства доступа в сеть поступают на сервер сети, где в соответствующих программных модулях подвергаются обработке сервером, согласно атрибутам аудиограммы удаленного абонента (для речевого сигнала ближнего абонента) и атрибутам аудиограммы ближнего абонента (для речевого сигнала удаленного абонента), которые выбраны из базы атрибутов в соответствии с номерами удаленного абонента и ближнего абонента. Далее обработанные сигналы через соответствующие устройства доступа в сеть передаются по сети связи к телефонным аппаратам абонентов.

Достоинством данного способа компенсации потери слуха в системах телефонии является возможность формирования персонализируемого аудиосигнала для слабослышащих пользователей на основе обработки речевого сигнала пользователя на сервере сети согласно атрибутам слабослышащего пользователя сети, которые хранятся в базе атрибутов на сервере сети связи и

доступны по номеру его телефона. При этом слабослышащий пользователь не использует свой СА во время телефонного разговора. По окончании разговора слабослышащий пользователь может снова использовать свой СА, что вызывает определенные неудобства для человека. СА у данного пользователя является основным инструментом в его активной жизнедеятельности. В тоже время слабослышащий человек со слуховым аппаратом испытывает ряд неудобств, обусловленных акустикой помещения, например, при восприятии звука от различных мультимедийных устройств, таких как музыкальные проигрыватели аудиосигнала, телевизоры и т.п.

Следует отметить, что этот способ не функционирует в цифровых телефонных сетях. Например, в сотовой телефонной сети требуется введение дополнительного декодирования/кодирования аудиосигнала на сервере сети, чтобы получить сигнал в виде импульсно-кодовой модуляции (ИКМ) для выполнения обработки сигналов ближнего и удаленного пользователей [62-64].

Обработка сигналов пользователей в данном способе на сервере сети связи ведется в широкополосном частотном диапазоне на основе функции обратной к частотной характеристике слуха слабослышащего пользователя, а также для компенсации потери слуха может включаться дополнительное усиление, применяемое к широкополосному сигналу и ограничитель мощности выходного сигнала, чтобы аудиосигнал всегда сохранял умеренную громкость даже, если пользователь на другом конце сети начинает говорить очень громко.

Однако у слабослышащих людей наблюдается утрата частотной избирательности - для чего необходимо обрабатывать аудиосигнал, в соответствии с психо-акустической шкалой и необходимо увеличивать отношение сигнал-шум в принимаемом ТА аудиосигнале для одинакового уровня разборчивости речи с нормально слышащим человеком. Разборчивость речи выше, если частотное разрешение компрессора динамического диапазона максимально согласуется с частотным разрешением восприятия акустической информации человеком - шкалой Барков [66].

Таким образом, известный способ обладает следующими недостатками:

- искажения аудиосигнала и низкая разборчивость речи;
- невозможность формирования персонализированного аудиосигнала, позволяющего пользователю возможность прослушивания аудиофайлов, радио и т.д.;
- отсутствие возможности у пользователя получить персонализированный аудиосигнал, необходимый ему во время личного общения с рядом находящимся собеседником при одновременном приеме аудиосигнала из динамиков различных устройств;
- неосуществимость одновременного ведения слабослышащим пользователем переговоров с удаленным абонентом сети связи, с находящимся рядом собеседником и прослушивания аудиосигналов с мультимедийных устройств;
- невозможность использования в качестве сервера сети, кроме автоматической телефонной станции (АТС), компьютерных устройств в минимальной комплектации, состоящих из процессора, оперативного запоминающего устройства, устройства долговременного хранения информации и устройства, обеспечивающего доступ в сеть связи;
- отсутствие у пользователя данного устройства в соответствии с его предпочтениями реализации различных режимов: телефонного разговора, личного общения с собеседником, приема аудиосигнала из динамиков разнообразных устройств.

Для расширения функциональных возможностей, повышения качества звука и разборчивости речи в мобильных телефонных аппаратах и системах связи для слабослышащих абонентов в описанном выше способе компенсации потери слуха в телефонной системе, заключающемся в том, что формируют персонализированные аудиосигналы для слабослышащих пользователей на основе их атрибутов, полученных из аудиограмм - частотных характеристик слуха слабослышащего пользователя, хранящихся в базе данных на сервере сети

связи и имеющих привязку к телефонным номерам слабослышащих пользователей.

При этом на сервере связи обрабатывают аудиосигналы в широкополосном частотном диапазоне на основе атрибутов слабослышащего пользователя, регулируют мощность обработанных аудиосигналов в соответствии с атрибутами слабослышащего пользователя, передают отрегулированные и персонализированные аудиосигналы с сервера связи к телефонным аппаратам слабослышащих пользователей. Предлагается в качестве сети связи использовать сотовую сеть, а в качестве телефонного аппарата – мобильный телефонный аппарат, при этом осуществляя режим, совмещающий функции мобильного телефонного и слухового аппарата.

Возможны дополнительные варианты осуществления способа, в которых целесообразно, чтобы:

- для осуществления режима, совмещающего функции мобильного телефонного и слухового аппарата:

- инсталлируют (с помощью электронного носителя информации или с персонального компьютера при помощи сети Интернет) в мобильный телефонный аппарат слабослышащего пользователя со встроенным беспроводным каналом связи программные модули - динамической компрессии аудиосигнала на основе атрибутов слуха пользователя и компенсации акустической обратной связи;

- микшируют сигнал микрофона мобильного телефонного аппарата для человека-собеседника, находящегося рядом со слабослышащим пользователем, с аудиосигналом, принятым по каналу беспроводной связи от мультимедийного устройства;

- производят динамическую компрессию микшированного аудиосигнала модулем динамической компрессии и компенсацию акустической обратной связи модулем компенсации акустической обратной связи;

- получают широкополосный аудиосигнал, который передают для воспроизведения мобильным телефонным аппаратом слабослышащего пользователя, при телефонном вызове мобильного телефонного аппарата слабослышащего пользователя;

- битовый поток сигнала оборудования оператора сотовой сети передают по привязанным телефонным номерам на сервер сети связи, который преобразует битовый поток сигнала оборудования оператора сотовой сети в сигнал импульсно-кодовой модуляции;

- по этому сигналу импульсно-кодовой модуляции формируют персонализированный аудиосигнал для слабослышащего пользователя на основе его атрибутов,

- на сервере сети связи персонализированный аудиосигнал кодируют и формируют битовый поток сигнала для данного персонализированного аудиосигнала, который передают по сети связи на мобильный телефонный аппарат слабослышащего пользователя для воспроизведения;

- на сервере связи дополнительно производили динамическую компрессию. При динамической компрессии формировали комплект субполосных аудиосигналов и управляли динамическим уровнем каждого субполосного аудиосигнала в каждой отдельной неравномерной субполосе частот в соответствии с частотной характеристикой слуха слабослышащего пользователя, коэффициентами алгоритма редактирования шума окружающей среды и функции компрессии динамического диапазона в отдельных неравномерных субполосах;

- при компенсации акустической обратной связи микшированный аудиосигнал дополнительно микшировали с выходным сигналом модуля компенсации акустической обратной связи, входным сигналом которого служит восстановленный широкополосный аудиосигнал модуля динамической компрессии. При этом расщепляют микшированный аудиосигнал и выходной сигнал модуля динамической компрессии на отдельные частотные каналы,

оценивают коэффициенты адаптивной фильтрации в каждом отдельном частотном канале, осуществляют адаптивную фильтрацию, сигнал которой служит выходным сигналом модуля компенсации акустической обратной связи.

Достоинством предложенного решения компенсации потери слуха и обеспечения доступа слабослышащих абонентов к ресурсам сети с помощью мобильных телефонов является совмещение функции телефона на мобильном телефоне с функцией слухового аппарата. В отличие от известных технических решений, наряду с коэффициентами динамической компрессии аудиосигнала, введена функция редактирования шумов окружающей среды, подавления обратной акустической связи.

Это дает возможность, благодаря функции формирования персонализированного аудиосигнала для слабослышащих пользователей сети, реализованной на сервере сети связи без использования СА и дополнительных устройств, увеличивать разборчивость речи слабослышащему пользователю, комфортно общаться с собеседником даже в неблагоприятной акустической обстановке (ресторанах, вокзалах), исключить «свист» обратной связи, мобильно переключаться на телефонный разговор.

Наличие беспроводного канала связи на мобильном ТА и функции СА дает возможность тугоухому человеку с высокой разборчивостью принимать сигнал аудиосигнал телевизора, сохранить качество звука музыкального центра и т.д., исключив шумовую обстановку среды.

Способ компенсации потери слуха в телефонной системе и в мобильном телефонном аппарате предлагается осуществлять с помощью разработанного устройства, функциональная схема которого представлена на рисунке 17.

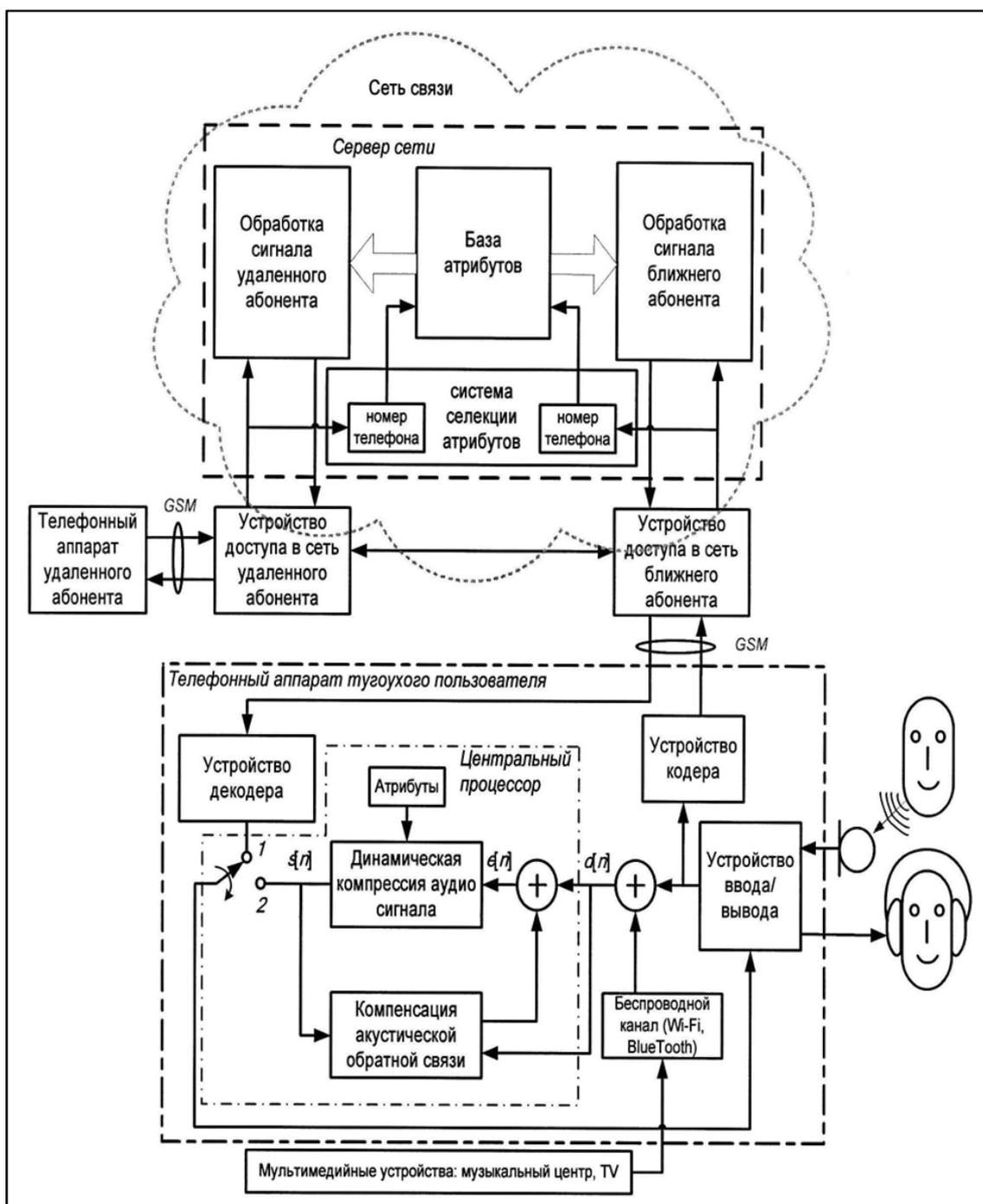


Рисунок 17. Функциональная схема системы связи, с помощью которой осуществляется реализация разработанного способа

Способ реализуется в сети связи, состоящей из ТА ближнего абонента и удаленного абонента - мобильного телефонного аппарата (МТА) слабослышащего пользователя, устройств доступа в сеть данных ТА, а также сервера сети связи, на котором размещается база атрибутов слабослышащих пользователей, программное обеспечение обработки сигналов ближнего и

удаленного абонентов и система селекции атрибутов по номеру тугухого абонента. Под МТА понимается любое программируемое персональное коммуникационное устройство, например смартфон, iPhone, iPad, а под телефонными номерами - любые идентификационные признаки пользователя, например, принятые в голосовой связи по протоколу IP-«Skype» и др.

Для работы МТА по первому варианту - в режиме слухового аппарата (СА) в существующие МТА со встроенным беспроводным каналом связи устанавливаются с помощью электронного носителя информации или с персонального компьютера при помощи сети Интернет программный модуль (программное обеспечение) динамической компрессии аудио сигнала на основе атрибутов слуха пользователя, полученные на основе аудиограмм слабослышащего пользователя, и модуль компенсации акустической обратной связи.

При работе в режиме СА коммутатор находится в положении 2 (рисунок 17). МТА включается, подключается беспроводной канал связи для прослушивания мультимедийных устройств (музыкальный центр, телевизор и т.п.). Сигнал беспроводного канала связи поступает на вход модуля динамической компрессии аудиосигнала, подвергаясь двойному микшированию (программными средствами).

На первое средство микширования из микрофона МТА через устройство ввода/вывода, выполненное на базе аналого-цифрового преобразователя (АЦП) и цифроаналогового преобразователя, поступают окружающие шумы. С выхода первого средства микширования сигнал $d[n]$ поступает на первый вход модуля компенсации акустической обратной связи (АОС) и на первый вход второго средства микширования, на второй вход которого поступает сигнал с выхода модуля компенсации АОС.

Сигнал $e[n]$ после второго микширования поступает на вход модуля динамической компрессии аудиосигнала, в котором подвергается сужению динамического диапазона в соответствии атрибутами (аудиограммой)

слабослышащего пользователя. С выхода модуля динамической компрессии аудиосигнала восстановленный сигнал $s[n]$ (блок восстановления находится на выходе модуля динамической компрессии и служит для восстановления широкополосности, он на рисунке 17 не показан) поступает на второй вход модуля компенсации АОС и на вход устройства ввода/вывода для его воспроизведения громкоговорителем МТА (преимущественно наушниками слабослышащего пользователя). Модуль компенсации акустической обратной связи выполнен на базе двух банков фильтров анализа АОС, банка фильтров синтеза АОС, блока субполосной обработки сигналов и предназначен для подавления акустической обратной связи.

Для работы МТА по второму варианту - в режиме слухового аппарата (СА) человек-собеседник, находящийся рядом со слабослышащим пользователем, разговаривает с ним. Аудиосигнал (речевой) с выхода микрофона через устройство ввода/вывода поступает вместе с окружающими шумами на первое средство микширования, поэтому основной входной аудиосигнал для модуля динамической компрессии формируется путем микширования сигнала микрофона с аудиосигналом, принятым из канала беспроводной связи от мультимедийного устройства.

Далее функционирование осуществляется согласно первому варианту. Слабослышащий пользователь, не прерываясь, может прослушивать как фразы собеседника, так и музыку, например, звучащую из музыкального центра. Если поступает телефонный звонок слабослышащий пользователь МТА переключает коммутатор в положение 1 (рисунок 17), МТА подключается к сети сотовой связи и находится в режиме функционирования телефона. Учитывая, что разработчики операционной системы МТА типа iPhone не дают возможности доступа к GSM кодеру-декодеру (преимущественно это обусловлено целями безопасности), битовый поток (например, GSM) канала перехватывается по телефонному номеру сервером сети связи (рисунок 17) у оборудования оператора мобильной связи (услуга сотового оператора).

Сервер сети связи преобразует битовый поток сигнала оборудования оператора сотовой сети в сигнал импульсно-кодовой модуляции (ИКМ). Дальнейшая обработка сигнала ИКМ производится в соответствии с находящейся на сервере программой формирования персонализированного аудиосигнала для слабослышащих пользователей на основе их атрибутов, полученных из аудиограмм, хранящихся в базе данных на сервере сети связи и имеющих привязку к телефонным номерам слабослышащих абонентов.

Сервер связи обрабатывает аудиосигналы в широкополосном частотном диапазоне на основе функции обратной к частотной характеристике слуха слабослышащего пользователя, усиливает и/или ограничивает мощность обработанных аудиосигналов в соответствии с функцией обратной к частотной характеристике слуха слабослышащего пользователя для сохранения умеренной громкости.

После обработки сервером связи формируется ИКМ код сигнала с учетом патологии слабослышащего пользователя. Далее данный код кодируется GSM кодером и передается в устройство доступа в сеть, а далее МТА принимает данный битовый поток из канала сети связи (приемо-передатчик МТА для простоты не показан), декодирует его устройством декодера. Декодированный сигнал поступает на вход устройства ввода/вывода и речевой аудиосигнал воспроизводится громкоговорителем МТА (наушниками). В зависимости от того, кто из абонентов: ближний или удаленный – является слабослышащим, возможны варианты способа, реализующего телефонный режим.

При работе МТА по третьему варианту - в режиме телефонного аппарата, удаленный абонент имеет нормальный слух, а ближний абонент - слабослышащий. В этом случае речевой аудиосигнал в обычном режиме передается через устройство кодера МТА в устройство доступа в сеть ближнего абонента, минуя сервер связи, и передается по сети связи при помощи оборудования оператора сотовой связи через устройство доступа в сеть удаленного абонента к ТА удаленного абонента.

Аудиосигнал удаленного абонента через устройство доступа в сеть на основании номера телефона слабослышащего абонента (ближнего абонента) поступает на сервер сети. На сервере связи сигнал удаленного абонента подвергается динамической компрессии согласно атрибутам аудиограммы ближнего абонента, которая выбирается из базы атрибутов в соответствии с номером телефона ближнего абонента. Далее обработанный и восстановленный сигнал удаленного абонента через устройство доступа в сеть передается по сети связи в устройство доступа в сеть ближнего абонента.

Как было описано выше, МТА ближнего абонента принимает данный битовый поток из канала сети связи, декодирует его устройством декодера. Декодированный сигнал поступает на вход устройства ввода/вывода и речевой аудиосигнал удаленного абонента воспроизводится громкоговорителем МТА (наушниками).

При работе МТА по четвертому варианту - в режиме телефонного аппарата, удаленный и ближний абоненты оба слабослышащие. В этом случае речевые сигналы ближнего и удаленного абонентов через соответствующие устройства доступа в сеть поступают на сервер сети связи, где подвергаются динамической компрессии согласно атрибутам аудиограммы удаленного абонента (для речевого сигнала ближнего абонента) и атрибутам аудиограммы ближнего абонента (для речевого сигнала удаленного абонента), которые выбраны из базы атрибутов в соответствии с телефонными номерами удаленного абонента и ближнего абонента. Далее обработанные сигналы, восстановленные через соответствующие устройства доступа в сеть, передаются по сети связи к МТА обоих абонентов.

При работе МТА по пятому варианту - в режиме телефонного разговора, общения с рядом находящимся собеседником, прослушивания аудиосигнала из мультимедийных внешних устройств и принятого от мультимедийного программного обеспечения, находящегося в МТА и предназначенного для воспроизведения аудиофайлов, радио и т.п., пользователь переключает

коммутатор в одновременное положение 1 и 2. При этом одновременно реализуются все предыдущие четыре варианта режимов. Поэтому пользователь имеет возможность получить персонализированный аудиосигнал при одновременном общении пользователя с другим абонентом по телефону и с собеседником лично, при приеме аудио сигнала из динамиков различных устройств, например во время просмотра передач телевидения, прослушивания музыки и т.п. Переключением коммутатора слабослышащий пользователь имеет возможность управлять режимами: телефонного разговора, личного общения с собеседником, приема аудиосигнала из динамиков и мультимедийных устройств. Для формирования персонализированного сигнала слабослышащего пользователя центральный процессор его МТА функционирует следующим образом (рисунок 18).

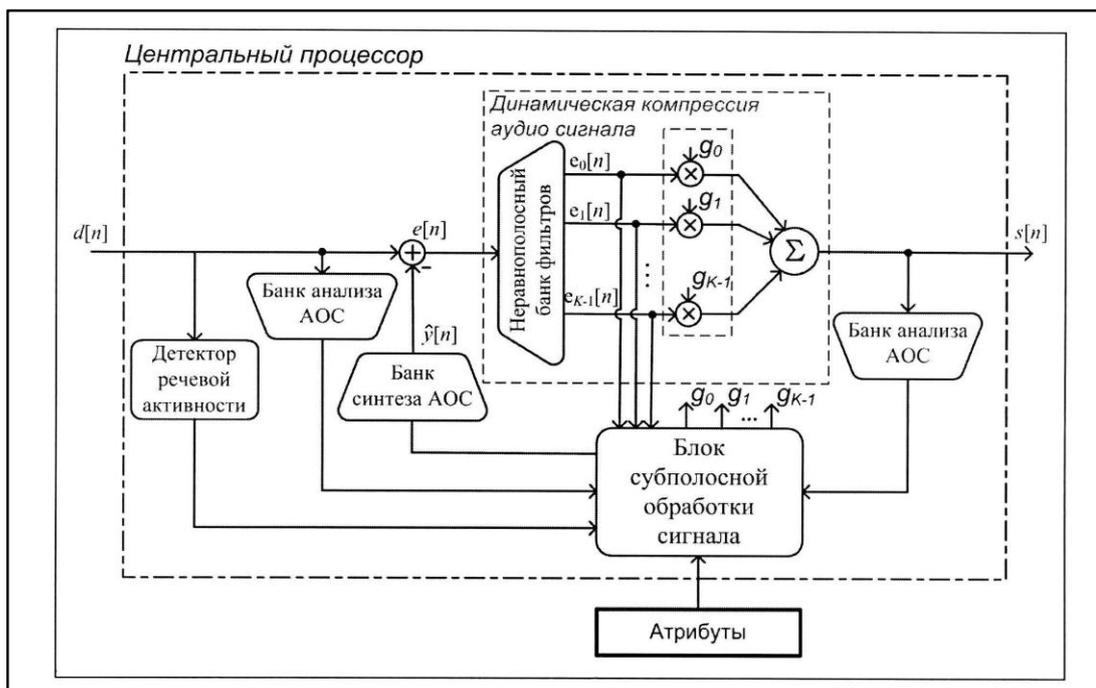


Рисунок 18. Функциональная схема формирования центральным процессором персонализированного аудиосигнала для слабослышащих пользователей сети

Формирование персонализированного аудиосигнала на центральном процессоре МТА осуществляется с помощью программного обеспечения динамической компрессии аудио сигналов. Программный блок состоит из неравнополосного банка фильтров, канальных умножителей на корректирующие

коэффициенты усиления, выходного сумматора восстановления широкополосности сигнала.

Компенсация акустической обратной связи производится на основе субполосной адаптивной фильтрации, в программный модуль которой включены два банка фильтров анализа АОС, банк фильтров синтеза АОС, (для краткости показаны на рисунке 18, как блок анализа АОС и банк синтеза АОС), блок субполосной обработки сигнала, который оценивает и обновляет коэффициенты адаптивной фильтрации, измеряет спектральную плотность мощности шума на основе вероятностной оценки наличия речевой паузы, определенной программой детектора речевой активности, и рассчитывает весовые коэффициенты алгоритма редактирования шума окружающей среды.

Сигнал $d[n]$ (на рисунках 17-18) поступает с выхода первого средства микширования на вход детектора речевой активности, на вход первого банка фильтров анализа АОС и на первый вход второго средства микширования, на второй вход которого поступает сигнал $\hat{y}[n]$ с первого выхода банка фильтров синтеза АОС.

Сигнал $e[n]$ с выхода второго средства микширования поступает на вход неравнополосного банка фильтров. Сигналы с выходов детектора речевой активности, первого банка анализа АОС, соответственно поступают на первый и второй входы блока субполосной обработки сигнала. Неравнополосный банк фильтров имеет K выходов, на которых с каждого фильтра, входящего в состав банка, получают сигналы $e_0[n] \dots e_{K-1}[n]$. Эти сигналы поступают на соответствующие входы данных блока субполосной обработки сигнала. Блок субполосной обработки сигнала производит расчет субполосных коэффициентов усиления $g_0 \dots g_{K-1}$.

Сигналы $e_0[n] \dots e_{K-1}[n]$ и коэффициенты $g_0 \dots g_{K-1}$ соответственно поступают с выходов данных неравнополосного банка фильтров и с выходов данных блока субполосной обработки сигнала на первый и второй входы набора средств микширования, выходы которых подсоединены к входам

многовходового сумматора, служащего для восстановления широкополосности, с выхода которого получают сигнал $s[n]$ для его воспроизведения МТА слабослышащего пользователя.

Выход сумматора соединен со вторым блоком анализа АОС, выход которого подсоединен к третьему входу блока субполосной обработки сигнала. Выход блока субполосной обработки сигнала соединен с входом блока синтеза АОС. В блок субполосной обработки сигнала вводятся данные об атрибутах, соответствующих аудиограмме конкретного пользователя.

Сигнал $d[n]$ (на рисунках 17-18) поступает с выхода первого средства микширования на вход детектора речевой активности, на вход первого банка анализа АОС и на первый вход второго средства микширования, на второй вход которого поступает сигнал с первого выхода банка синтеза АОС. Сигнал $e[n]$ с выхода второго средства микширования поступает на вход неравнополосного банка фильтров.

Сигналы с выходов детектора речевой активности, первого банка анализа АОС, со второго выхода банка синтеза АОС соответственно поступают на первый, второй и третий входы блока субполосной обработки сигнала. Неравнополосный банк фильтров имеет K выходов, на которых с каждого фильтра, входящего в состав банка, получают сигналы $e_0[n] \dots e_{K-1}[n]$. Эти сигналы поступают на соответствующие входы данных блока субполосной обработки сигнала.

Блок субполосной обработки сигнала производит расчет субполосных коэффициентов усиления g_K . Отсчеты канальных сигналов $e_0[n] \dots e_{K-1}[n]$ и коэффициенты $g_0 \dots g_{K-1}$ соответственно поступают на первый и второй входы канальных умножителей на корректирующие коэффициенты усиления, выходы которых подсоединены к входам многовходового сумматора, служащего для восстановления широкополосности, с выхода которого получают сигнал $s[n]$ для его воспроизведения МТА слабослышащего пользователя.

Выход сумматора соединен со входом второго блока анализа АОС, выход которого подсоединен к третьему входу блока субполосной обработки сигнала. В блок субполосной обработки сигнала вводятся данные об атрибутах, соответствующих аудиограмме конкретного пользователя.

Блок субполосной обработки управляет уровнем сигнала в соответствующих субполосах для обеспечения требуемой динамики уровней субполосных сигналов, обусловленных аудиограммой (атрибутами) слабослышащего пользователя МГА, коэффициентами алгоритма редактирования шума окружающей среды и функции компрессии динамического диапазона в соответствующих субполосах, которые интегрируются в соответствующие коэффициенты субполосного усиления g_k . Компрессия динамического диапазона (КДД) используется для уменьшения разницы в уровнях компонент с большой и малой интенсивностью в аудиосигнале. При этом широкий динамический диапазон речевого сигнала отображается в суженный динамический диапазон остаточного слуха.

В предлагаемом способе в качестве неравнополосного банка фильтров использовалась схема фильтрации с малой (менее 4 мс) групповой задержкой на основе кохлеарного банка фильтров, который реализован в виде набора параллельных полосовых фильтров с бесконечной импульсной характеристикой (БИХ) второго порядка. Кохлеарный банк фильтров обладает несколькими важными желательными свойствами:

- декомпозиция сигнала выполняется на критические полосы слуховой системы человека;
- маленькая (менее 4 мс) групповая задержка;
- высокая вычислительная эффективность (фильтрация в каждом канале выполняется при помощи БИХ фильтра второго порядка).

В техническом решении использовался 22-канальный банк фильтров, основанный на разностной кохлеарной модели второго порядка. В соответствии с имеющимися пороговыми значениями по атрибутам слабослышащего

пользователя блок субполосной обработки сигнала вычисляет корректирующие коэффициенты усиления $g_0 \dots g_{K-1}$ сигнала в каждой субполосе. Далее используется алгоритм компрессии, поскольку динамический диапазон выходного сигнала ограничен болевым порогом. Основная идея алгоритма компрессии динамического диапазона (КДД) заключается в автоматическом управлении коэффициентами усиления в зависимости от текущего уровня входного сигнала. Основными параметрами КДД являются функция ввода/вывода и время атаки и восстановления.

Сигналы в субполосах с высокой мощностью ослабляются, а с малой мощностью усиливаются. Благодаря такой обработке тихие звуки становятся слышимыми, а громкие звуки не вызывают дискомфортных ощущений. Таким образом, КДД заключается в автоматическом управлении коэффициентами усиления в зависимости от текущего уровня входного сигнала.

Основными параметрами КДД являются следующие: порог компрессии (СТ); коэффициент сжатия (CR); время атаки и отпущания; коэффициент усиления слухового аппарата (G_{dB}). Порог компрессии (СТ), измеряемый в децибелах, определяет точку изгиба характеристики вход/выход компрессора, после которой алгоритм КДД становится активным. Если уровень входного сигнала ниже СТ, то выходной сигнал усиливается линейно.

В случае, когда уровень выходного сигнала выше порога компрессии (СТ), коэффициент усиления компрессора уменьшается. Параметр CR определяет степень сжатия динамического диапазона. Например, значение CR=5 (или 5:1) означает, что на каждые 5 дБ увеличения уровня входного сигнала, уровень выходного сигнала возрастет только на 1 дБ. На рисунке 19 приводится характеристика вход/выход компрессора для параметров: CR=2, СТ=70 дБ и $G_{dB}=10$ дБ.

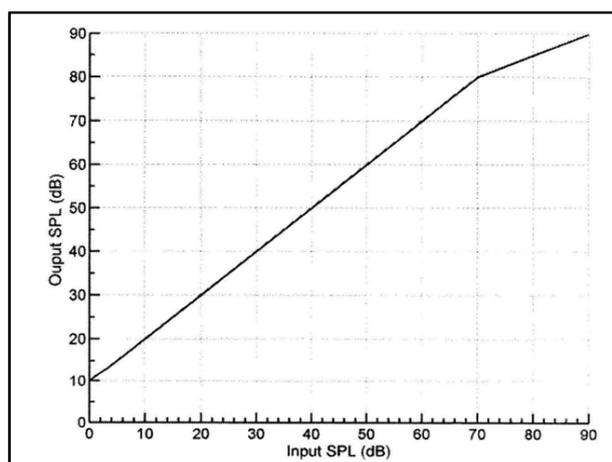


Рисунок 19. Характеристика вход/выход компрессора

Данный график определяет соотношение между входным и выходным уровнями звукового давления (sound pressure level - SPL) в компрессоре.

На рисунках 20-а и 20-б показан пример результата обработки входного сигнала, состоящего из двух частей – громкого (рисунок 20-а) и тихого участка алгоритмом КДД (рисунок 20-б).

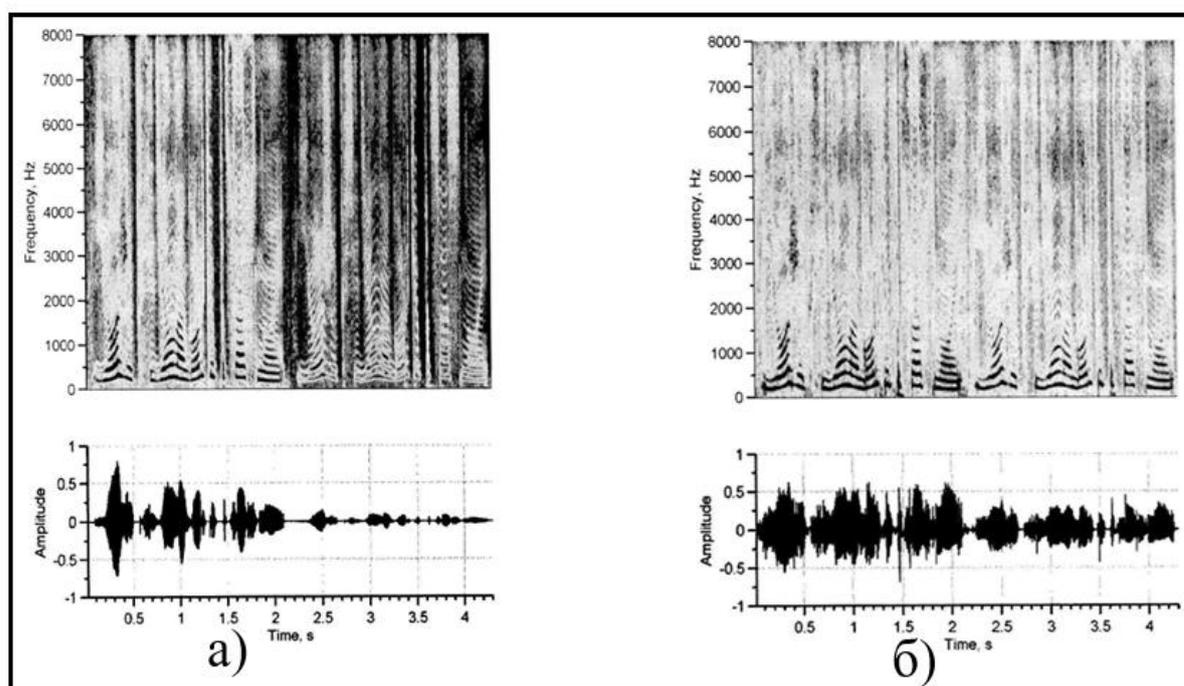


Рисунок 20. а) входной сигнал, б) входной сигнал, полученный в результате компрессии динамического диапазона (КДД)

Отчетливо виден эффект нелинейного усиления (обе части приблизительно выровнялись по громкости (рисунок 20-б). Видимые в спектре после обработки искажения появляются в результате нелинейной обработки в компрессоре, но не оказывают ощутимого влияния на разборчивость речи и узнаваемость диктора.

Тестовый речевой сигнал (рисунок 21-а) обрабатывался при помощи алгоритма КДД, настроенного на конкретный профиль потери слуха. Полученная после обработки сигнала спектральная гистограмма представлена на рисунке 21-б. Результаты показывают, что алгоритм КДД позволяет адаптировать уровень выходного сигнала к слуховой характеристике слабослышащего пользователя.

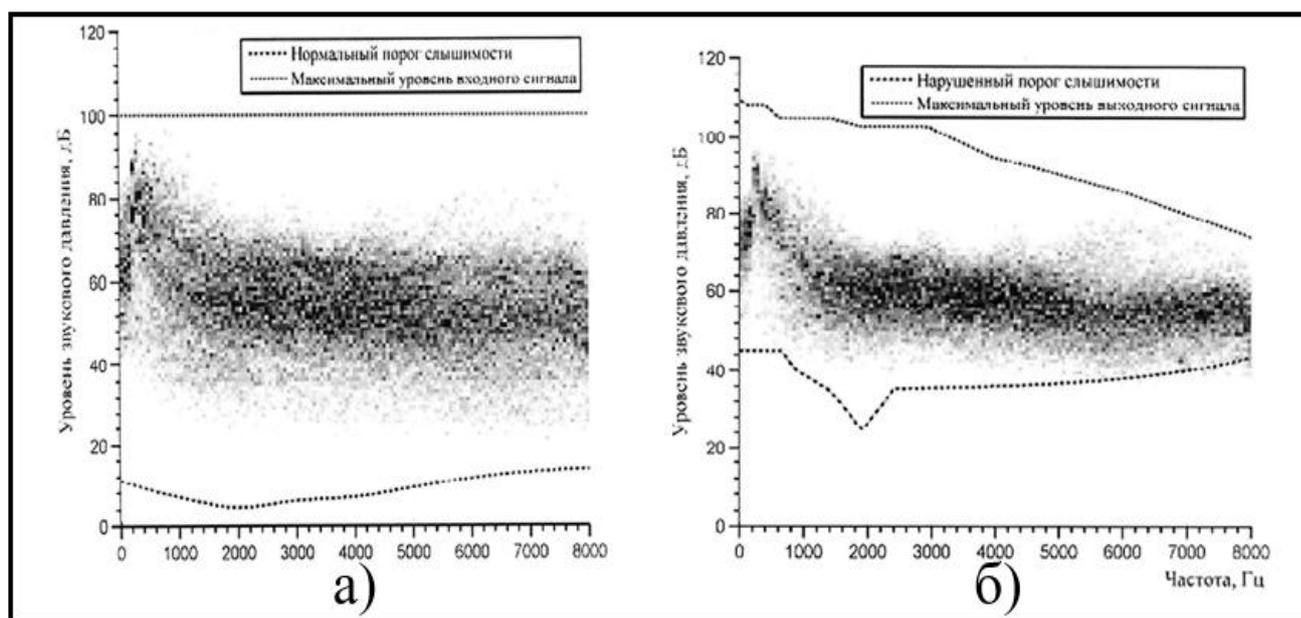


Рисунок 21 – а) спектральная гистограмма аудиосигнала (речевого), б) спектральная гистограмма сигнала, обработанного при помощи КДД

В основе используемого алгоритма редактирования шума окружающей среды лежит психо-акустически мотивированное правило спектрального взвешивания.

Алгоритм использует настраиваемый параметр $\zeta=10^{-RL/10}$, который определяет желаемый уровень остаточного шума RL в дБ. Оценка спектральной

плотности мощности шума (СПМ) выполняется для каждого канала алгоритма КДД при помощи вычислительно эффективного и устойчивого к ошибкам алгоритма на основе модифицированного метода MCRA (Minima Controlled Recursive Averaging).

Текущее значение СПМ шума R_n (где n - номер отсчета) вычисляется усреднением предыдущих значений СПМ $R_e(n)$, используя сглаживающие параметры, зависящие от вероятности наличия полезного сигнала, который определяет детектор речевой активности, например, кепстрального анализа. Параметры обновляются каждые 4 мс. На сервере связи также может быть произведена аналогичная динамическая компрессия, но без компенсации АОС и шумоподавления.

На рисунках 22-а и 22-б представлены результаты работы алгоритма редактирования шума окружающей среды: рисунок 22-а - сигнал на входе микрофона, а на рисунке 22-б - обработанный.

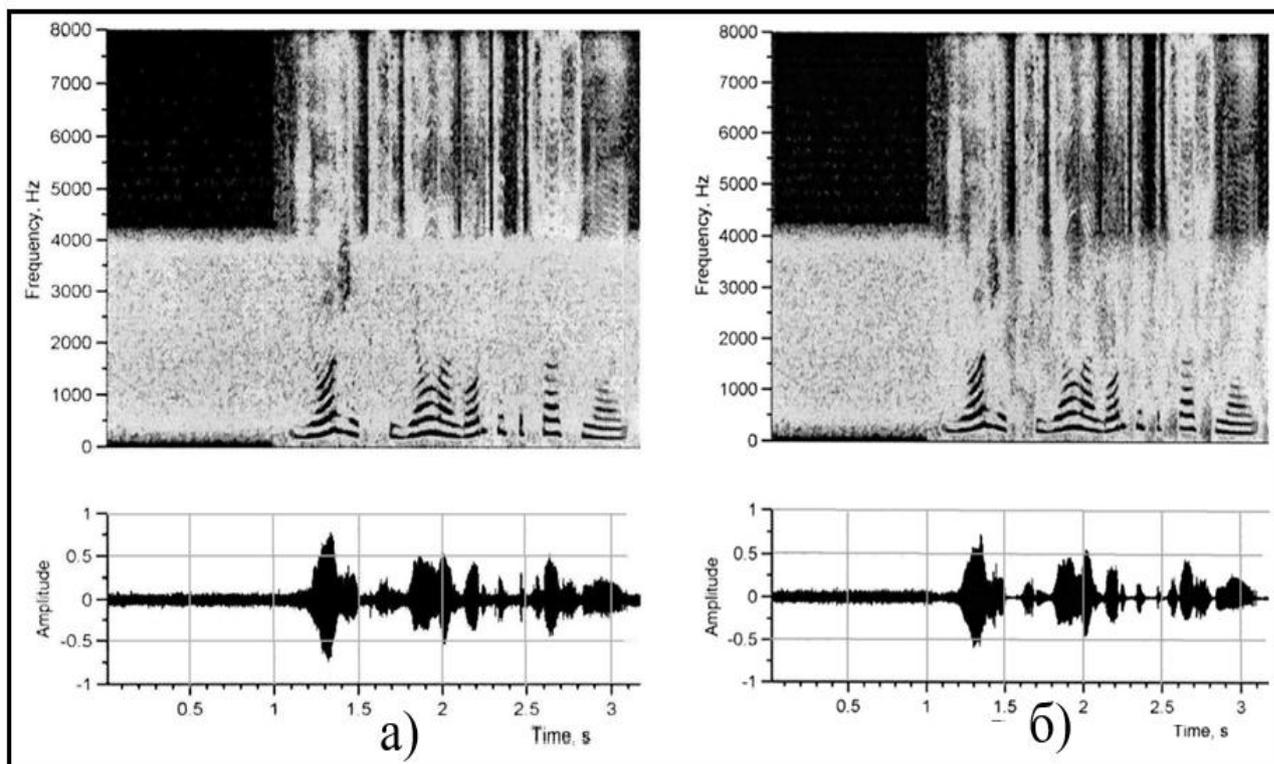


Рисунок 22. а) диаграмма входного аудиосигнала, б) после обработки алгоритмом редактирования шума

Подавление акустической обратной связи выполняется следующим образом (рисунок 23). При помощи первого банка фильтров анализа АОС выполняется расщепление сигнала $d[n]$ на входе КДД на M спектральных компонент. При помощи второго банка анализа АОС, идентичного первому, выполняется расщепление сигнала $s[n]$ выходе КДД на M спектральных компонент. Поскольку спектры сигналов внутри каналов занимают более узкие частотные полосы, выполняется переход на более низкую частоту дискретизации.

В банке фильтров синтеза АОС (банк фильтров) исходная частота дискретизации восстанавливается. Блок субполосной обработки сигналов (рисунок 18) в каждом канале банка фильтров оценивает свой вектор коэффициентов адаптивного фильтра. Последние результаты в области адаптивной фильтрации показывают, что неравнополосные адаптивные структуры превосходят равнополосные по таким параметрам как скорость сходимости и/или модельная ошибка, благодаря их повышенной гибкости. Для субполосной декомпозиции сигнала в техническом решении используется передискретизированный неравнополосный косинусно-модулированный банк фильтров (КМБФ), амплитудно-частотная характеристика которого показана на рисунке 23.

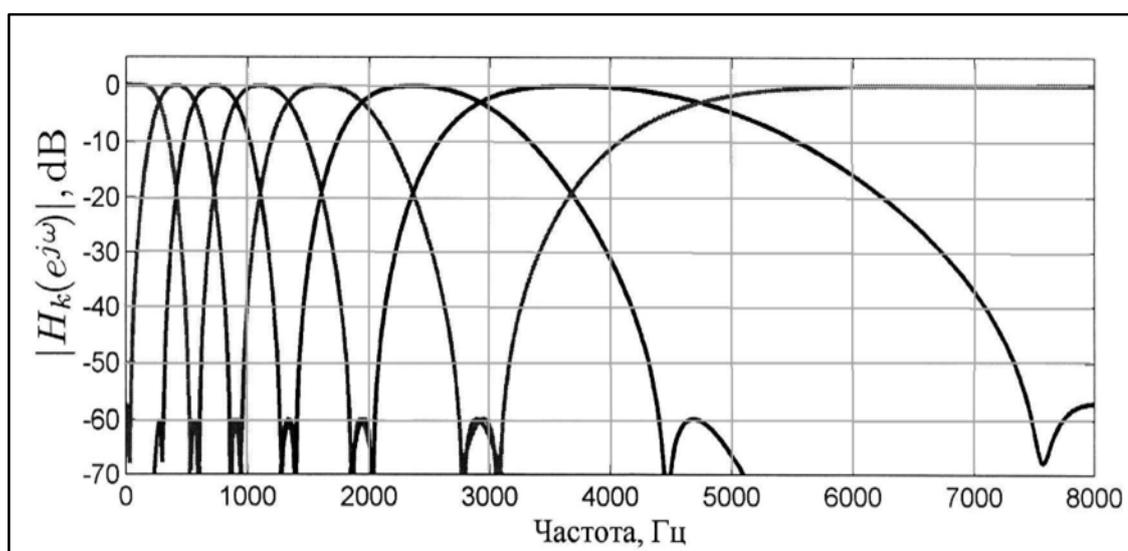


Рисунок 23. Амплитудно-частотная характеристика (АЧХ) банка фильтров анализа

В каждом канале оценивается свой набор коэффициентов адаптивного фильтра. Процедура оценки одинакова для всех каналов и отличается лишь значениями параметров, таких как порядок фильтра, коэффициент потерь и шаг адаптации. Обновление коэффициентов реализуется на основе алгоритма наименьших квадратов (для упрощения записи индекс номера канала опущен):

1. Каждому коэффициенту фильтра $w[l]$, $l=0, 1, \dots, L-1$ присваивается нулевое значение, L - порядок адаптивного фильтра.

2. Вычисляется выходной отсчет фильтра:

$$\hat{Y}[m] = \sum_{l=0}^{L-1} w[l]s[m-l], \quad (2.21)$$

где m - номер текущего отсчета входного сигнала,
 $s[m]$ - входной сигнал.

3. Вычисляется оценка ошибки:

$$e[m] = d[m] - \hat{Y}[m], \quad (2.22)$$

где $d[m]$ - желаемый сигнал.

4. Обновляются весовые коэффициенты:

$$w[l] = \delta w[l] + 2\mu e[m] \times [m-l], \quad (2.23)$$

где $0 < \delta < 1$ - коэффициент потерь.

Параметр μ , представляет собой шаг адаптации алгоритма.

5. Увеличивается номер текущего отсчета: $m=m+1$. Алгоритм переходит к шагу 2.

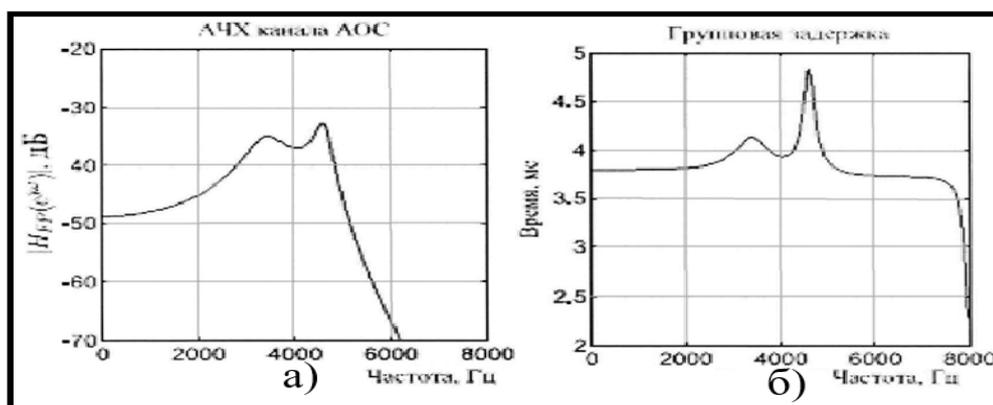


Рисунок 24. а) Амплитудно-частотная характеристика (АЧХ) канала акустической обратной связи (АОС);
 б) групповая задержка канала акустической обратной связи (АОС)

На рисунках 24-а и 24-б изображены частотные характеристики моделируемого канала акустической обратной связи.

Для моделирования прямого канала, т.е. обработки сигнала в МТА, выбрана усредненная АЧХ (рисунок 25), которая компенсирует типичное повреждение слуховой системы.

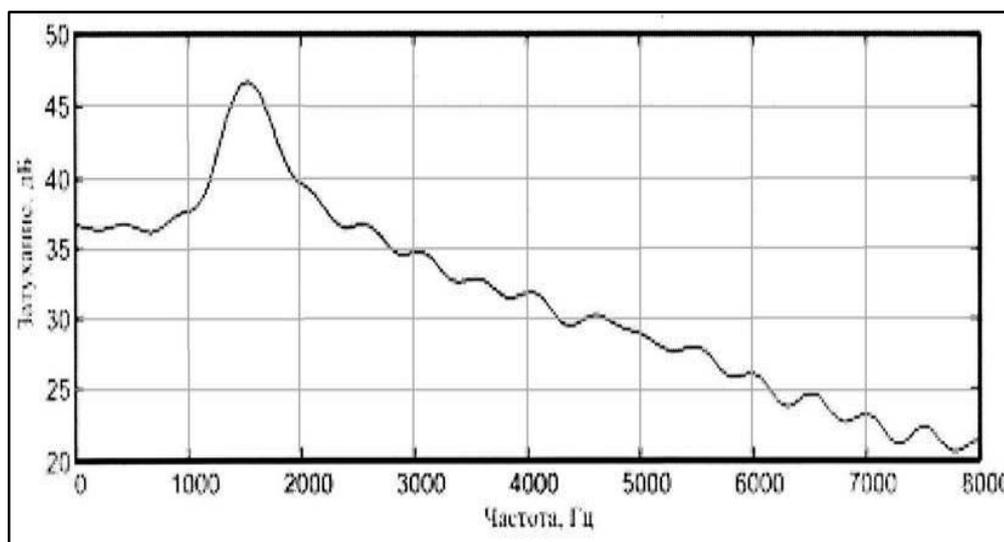


Рисунок 25. Модель частотной характеристики мобильного телефонного аппарата для реализации заявленного способа

Наибольшие потери происходят в районе 1,5 кГц, т.е. в том диапазоне частот, где речь имеет наибольшую информативность.

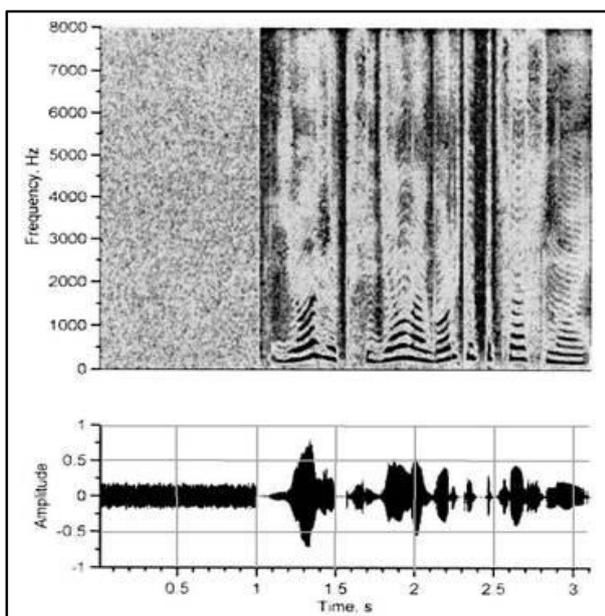


Рисунок 26. Входной аудиосигнал до подавления акустической обратной связи (АОС)

На рисунках 26, 27-а и 27-б показаны результаты работы модуля АОС:

- рисунок 26 - входной аудиосигнал,
- рисунок 27-а - аудиосигнал на выходе динамика, из характеристики отчетливо видно возбуждение системы на частоте около 5000 Гц,
- рисунок 27-б- результат обработки входного аудиосигнала алгоритмом подавления акустической обратной связи.

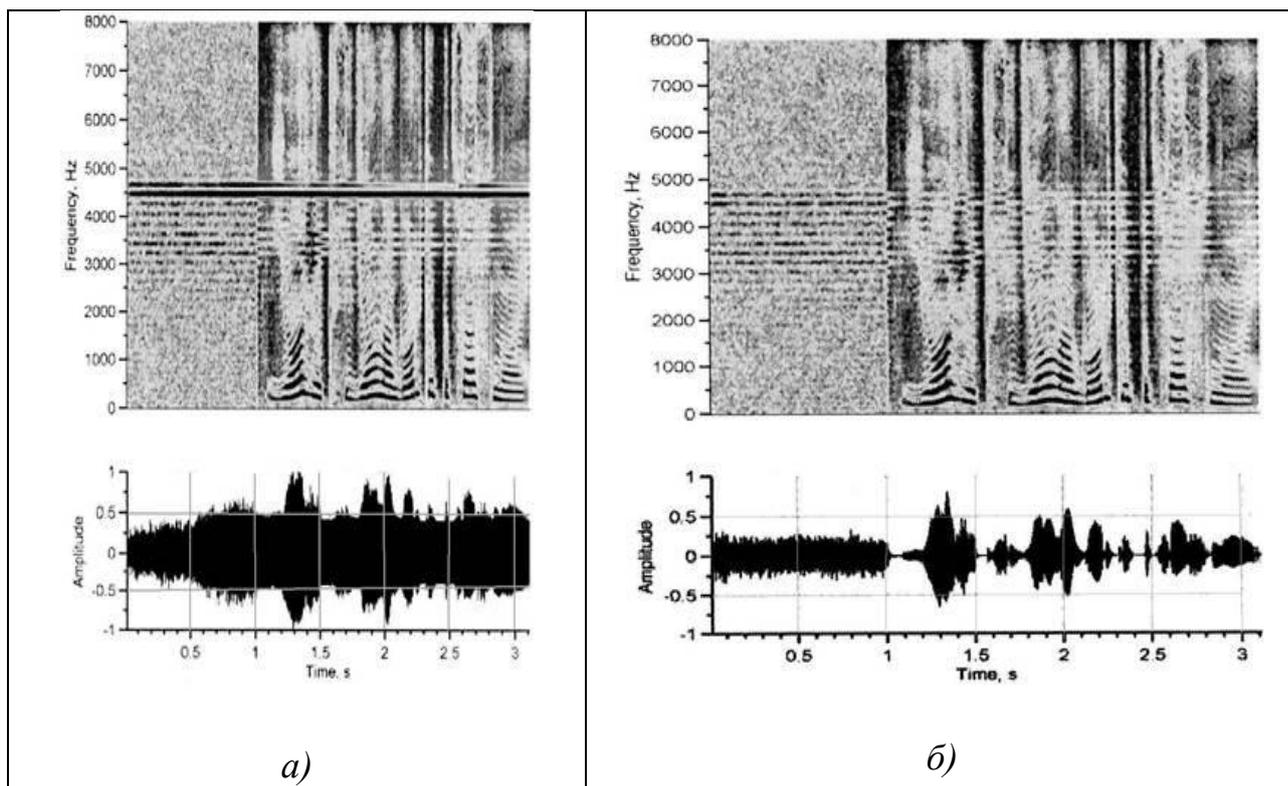


Рисунок 27. а) Аудиосигнал на выходе громкоговорителя мобильного телефонного аппарата (без АОС), б) аудиосигнал на выходе громкоговорителя мобильного телефонного аппарата, полученного в результате обработки входного аудиосигнала алгоритмом подавления акустической обратной связи (АОС)

Из приведенных спектрограмм видно, что использование алгоритма подавления АОС позволяет использовать более высокие коэффициенты усиления для обработки сигнала в прямом канале, что приводит к увеличению разборчивости речи слабослышащим пользователем.

2.3 ПРИКЛАДНАЯ РЕАЛИЗАЦИЯ УСТРОЙСТВ НА ОСНОВЕ ПРЕДЛОЖЕННЫХ МЕТОДОВ

В настоящее время существует ряд мобильных приложений, которые берут на себя функции слухового аппарата, такие как EarMachine, Jacoti и Mobile Ears [66]. Некоторые из подобных мобильных приложений, помимо коррекции слуха, предоставляют возможность дистанционной настройки и тестирования специалистом. Во время дистанционного сеанса может быть проведено:

- обследование состояния слуха;
- демонстрация функций коррекции слуха;
- настройка средства коррекции слуха в соответствии с результатами обследования;
- дополнительная персонализация настроек приложения;
- краткий инструктаж по использованию средства коррекции;
- введение в курс сурдологической адаптации для ускорения привыкания к звуку после коррекции и т.д.

Возможность удаленного общения со специалистом облегчает для пациента процесс настройки и эксплуатации средства коррекции, снимает когнитивный диссонанс, относительно правильности/верности настройки, позволяет сохранить и отслеживать в дальнейшем результаты обследования, экономит время, которое ему пришлось бы потратить на физическое посещение аудиолога. На рисунке 28 показано количество потенциальных пользователей подобных приложений для коррекции слуха.

Статистика говорит, что только 16% американцев в возрасте от 20 до 64, имеющих нарушения слуха, когда-либо использовали слуховые аппараты [35, 36]. Не более 31,6 млн. человек в США могут получить доступ к дистанционной аудиометрии, из них 8,7 млн. пожилых людей (65 лет и старше). В Китае распространенность СА среди слабослышащих составляет 6,5%. При этом, например, 80% владельцев слуховых аппаратов в Китае не используют их из-за психологического дискомфорта, вызванного стигматизацией слуховых

нарушений, или вследствие технического несовершенства устройств и сложности настройки [37]. Получить доступ к дистанционной аудиометрии могут не более 102,8 млн. человек, среди них 24,8 млн. пожилого возраста.

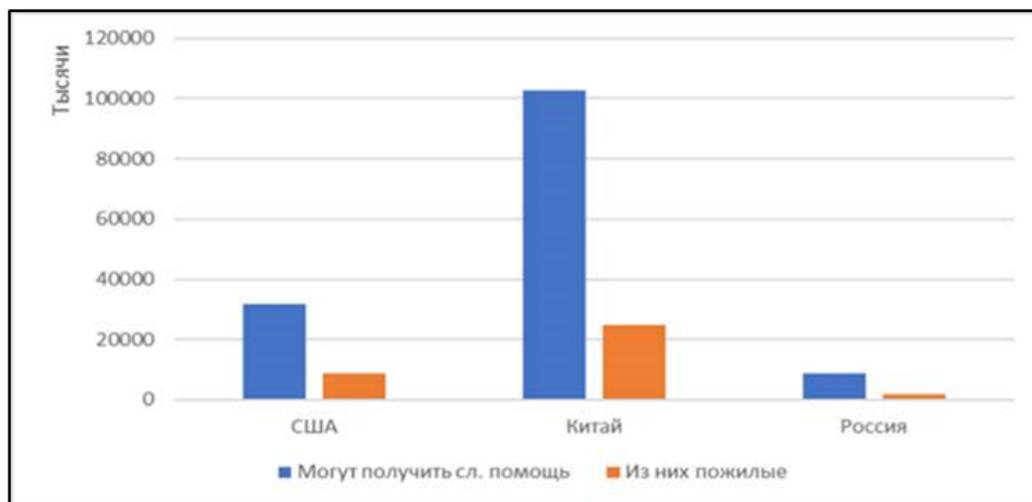


Рисунок 28. Потенциальная аудитория пользователей приложений для коррекции слуха

В России распространенность СА среди слабослышащих составляет 5,2% [38], получить доступ к дистанционной аудиометрии могут до 8,75 млн. человек, среди них 1,87 млн. пожилых.

Часть ведущих производителей СА, таких как Oticon (Дания), Phonak (США), ReSound (GN, международный концерн), Widex (Дания) предоставляют функции дистанционной настройки для ряда моделей. В таблице 1 представлена информация о доступных пользователю функциях приложений от ведущих производителей СА.

Так, компании Oticon и Phonak дают возможность настроить слуховой аппарат с помощью специалиста через мобильное приложение, однако обновление программного обеспечения (ПО) слуховых аппаратов выполняется исключительно при визите к специалисту.

Компании Widex и ReSound дают пользователю возможность обновить ПО, и связаться со специалистом службы поддержки через мобильное приложение.

Таблица 1. Возможности средств коррекции ведущих производителей СА

Производитель/ Функция	Oticon	Phonak	ReSound	Widex
Обновления ПО устройства	×	×	✓	✓
Возможность видеозвонка со специалистом	✓	✓	✓	✓
Удаленная настройка в режиме реального времени	✓	✓	✓	✓
Необходимые аксессуары	×	×	×	REMOTE LINK
Модели СА	Opn S™, Opn™, Xceed, Opn Play™, Xceed Play, and Siya	Audéo™ M, Bolero™ M, Virto M, Naida M and Audéo™ B-Direct	LiNX Quattro, LiNX 3D, Enzo Q, and Enzo 3D	MOMENT™, EVOKE™, BEYOND™, UNIQUE™, DREAM™

На основе разработанных и представленных в работе ранее методов и компьютерных устройств было разработано приложение Petralex для смартфонов под ОС Android и iOS, обеспечивающее компенсацию потери слуха в условиях агрессивной шумовой обстановки, позволяет получить доступ к системе дистанционного тестирования слуха и настройки мобильного средства [38].

Дистанционная настройка устройства коррекции слуха также важна для пользователей, ограниченных в возможности посещения специалиста-аудиолога. По данным исследования [71], существенной разницы в результатах настройки корректирующего устройства очно и через разработанные приложения для удаленной настройки не выявлено.

Разработанная система дистанционной настройки слухового аппарата (СДН СА) позволяет использовать ее и на мобильных платформах, и на персональных компьютерах. Система реализована на базе кроссплатформенной open-source технологии WebRTC (Web Real-Time Communication - «Коммуникации по сети в реальном времени») на открытых стандартах и API

(Application Programming Interface - программный интерфейс приложения), что упрощает ее интеграцию.

WebRTC поддерживает кодирование, декодирование и передачу данных между устройствами, используя стандартные (Opus и VP9) и дополнительные кодеки, которые используются для обеспечения совместимости. Технология WebRTC позволяет связать двух пользователей для видео- или аудио связи через Peer-to-Peer соединение. Для подключения пользователей, находящихся в сети с NAT (Network Address Translation - «Преобразование сетевых адресов»), необходимо использование STUN-серверов (Session Traversal Utilities for NAT - «Утилиты прохождения сессий для NAT») для определения общедоступного IP-адреса устройства. Когда пользователь вызывает специалиста, приложение отправляет SDP-пакет (Session Description Protocol - «Протокол описания сессии») через телекоммуникационный сервер с информацией о типе передаваемых данных и используемых кодеках, на основе полученного пакета генерируется ответный SDP-пакет и отправляется устройству пользователя, производится обмен информацией о настройках NAT и IP пользователей, и устанавливается P2P-соединение через протокол WebRTC, так же устанавливается WebSocket-соединение, через которое организуется общение в текстовом чате, а специалист получает доступ к переключению экранов и управлению рядом функций приложения пользователя. Поскольку данные о состоянии здоровья, в частности, слуха, относятся к наиболее чувствительному и защищаемому классу персональных данных, система предусматривает работу строго в P2P режиме, исключая доступ третьих лиц к результатам тестирования.

WebRTC использует ICE-протокол (Interactive Connectivity Establishment - «Установка интерактивного соединения») для определения наилучшего пути для передачи данных между устройствами через NAT и файрволы. ICE позволяет обходить препятствия, которые могут возникнуть при попытке установить прямое соединение между устройствами. WebRTC включает в себя механизмы шифрования данных и использует протоколы SRTP (Secure Real-time Transport

Protocol - «Безопасный протокол передачи данных в реальном времени») и DTLS (Datagram Transport Layer Security - «Протокол датаграмм безопасности транспортного уровня»), чтобы обеспечить безопасность передачи медиаданных между устройствами.

Тональный сигнал, который используется для проведения удаленного тестирования, генерируется на устройстве пользователя согласно отправленным параметрам (частота и амплитуда), которые определяются системой автоматически. Система автоматически распределяет пациентов генерируя цифровую очередь, и предоставляет инструменты для планирования и проведения сессий между специалистом и пациентом.

Принципиальная схема СДН СА представлена на рисунке 29.

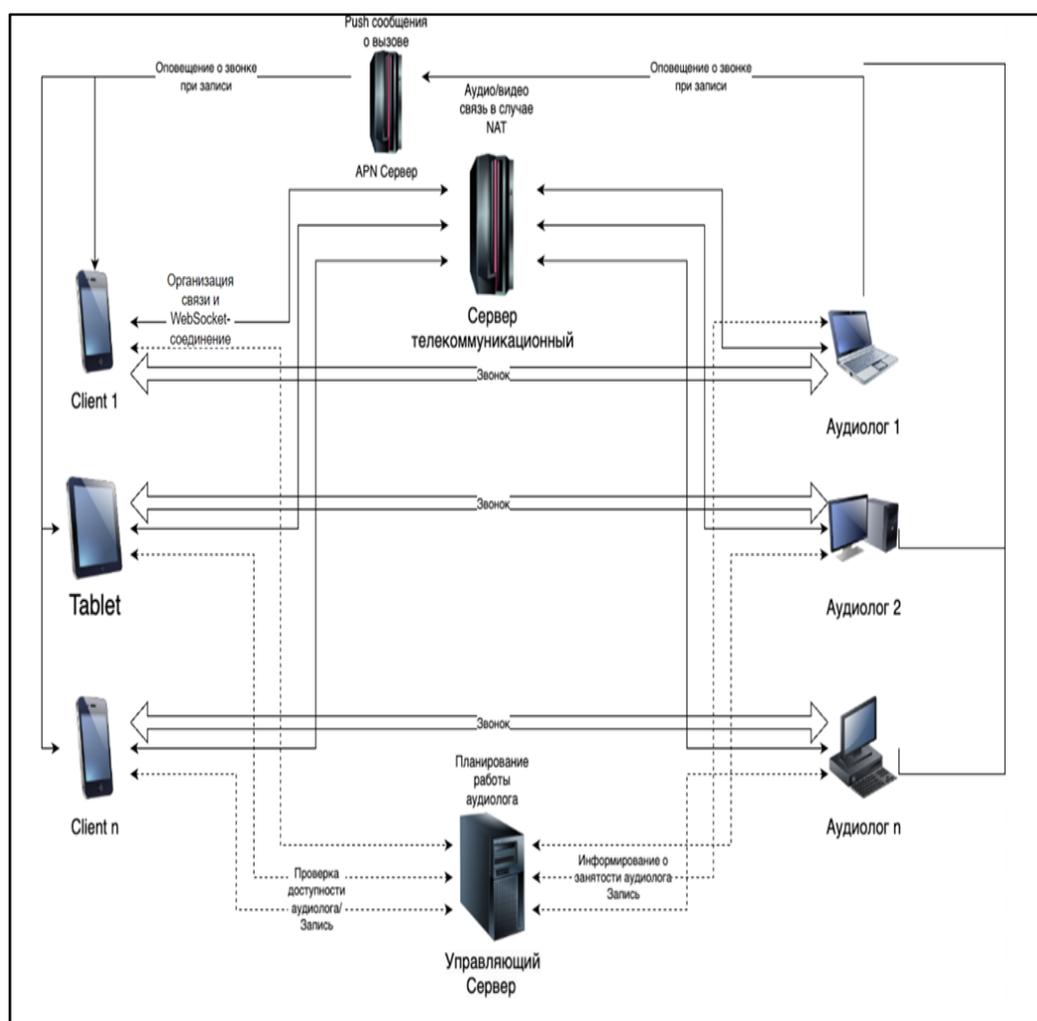


Рисунок 29. Система дистанционной настройки слухового аппарата

Система дистанционной настройки СА содержит следующие функциональные элементы (рисунок 29):

- Web интерфейс специалиста «Аудиолог 1, 2 ... n»,
- управляющий сервер, который отвечает за планирование работы специалиста,
- телекоммуникационный сервер - отвечает за организацию звонка,
- приложения на мобильном устройстве пользователя (платформы «iOS» и «Android»).

Рабочее место пользователя состоит из нескольких основных интерфейсов:

- экран голосового вызова,
- экран текстового чата,
- главный экран,
- экран настроек
- экран сервисов.

На главном экране приложения пользователь переключает существующие профили слуха пользователя (которые содержат информацию о частотах и соответствующим им степени усилении звука), добавляет профили, и может перейти на экран общения с консультантом. На экране общения с консультантом пользователь может выбрать одну из двух функций: начать звонок со специалистом или задать вопрос в текстовом чате. После установления соединения пользователь и специалист могут приступать к тестированию слуха [77-81]. Пользователь может изменить формат общения на аудио или текстовый чат. Во время взаимодействия с пользователем специалист контролирует процесс тестирования, демонстрирует, настраивает функции приложения, которые описаны в таблице 2.

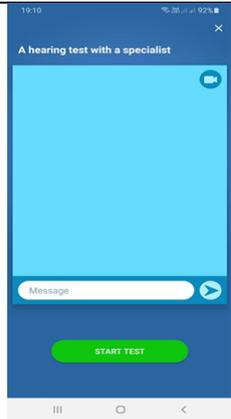
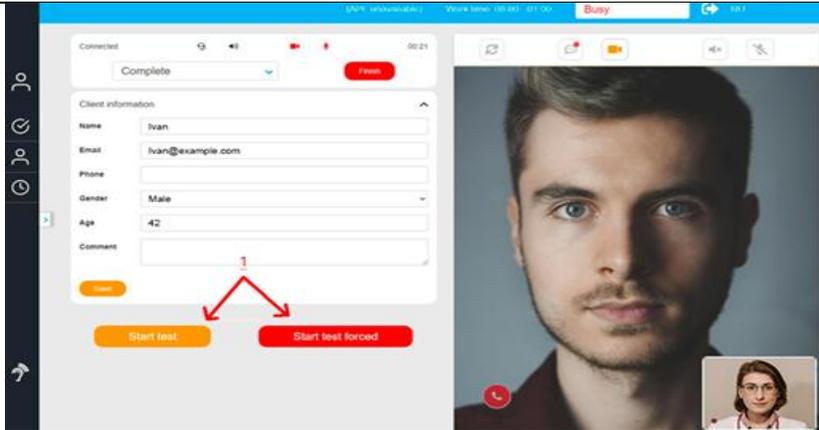
Специалист принимает вызовы в соответствии с расписанием. Календарь в кабинете специалиста отображает запланированные звонки с пользователями. При нажатии на дату календаря открывается список записей на эту дату.

Таблица 2. Функции специалиста по настройке

Функция	Описание функции
Усиление	Выбор формулы усиления на основе предпочтений пользователя (Super Boost, NAL, POGO, Berger, Petralex)
Усиление высоких частот	Регулирует усиления высоких частот
Усиление низких частот	Регулирует усиления низких частот
Баланс	Отношение громкости правого и левого динамика
Частотная компрессия	Сокращение динамического диапазона входного сигнала, уменьшение разницы между самыми тихими и громкими звуками
Шумоподавление	Подавление окружающих шумов во входном сигнале
Подавление акустической обратной связи (АОС)	Подавление эффекта акустической обратной связи (шум, который вызван образованием звуковой петли, вследствие попадания звукового сигнала динамика в микрофон)
Интеллектуальная адаптация к шуму	Обработка на основе динамического изменения порогов слухового восприятия в меняющейся шумовой обстановке

В списке звонков на текущую дату специалист может просмотреть информацию о каждом вызове (имя пользователя, время звонка и статус звонка).

Таблица 3. Графические интерфейсы: Запуск дистанционного тестирования

Интерфейс пользователя	Интерфейс специалиста
	

Когда специалист находится в статусе ожидания в личном кабинете, при вызове от пользователя специалист получает уведомление. Если звонок пользователя был заранее запланирован на определённое время, то за пять минут до начала звонка у специалиста есть возможность подключиться к звонку и

ожидать подключения пользователя. При необходимости специалист может отправить пользователю push-напоминание.

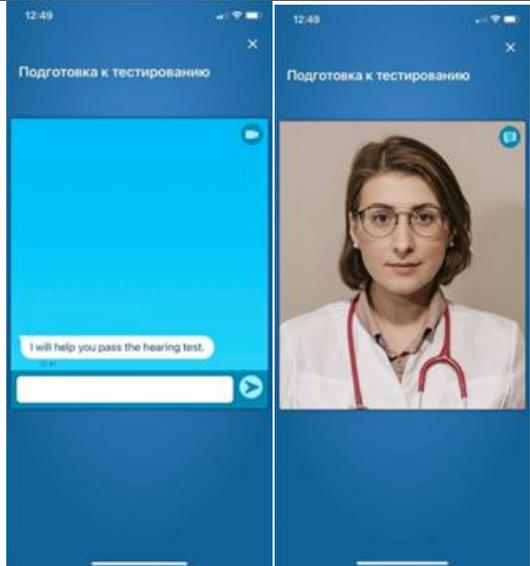
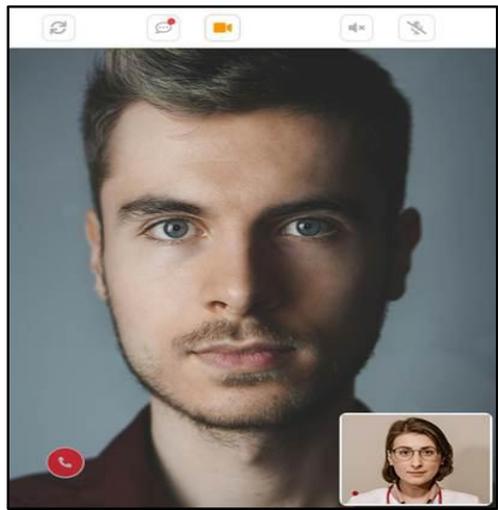
Если на момент запланированного звонка специалист все еще занят с другим пациентом, специалисту выдается предупреждение о параллельном звонке по записи. По окончании текущего сеанса специалист переходит к следующему.

Интерфейсы, с помощью которых пользователь и специалист взаимодействуют во время тестирования продемонстрированы в таблицах 3-7.

В интерфейсе предусмотрены следующие варианты начала тестирования (таблица 3):

- специалист нажимает кнопку «Start test forced» - тест начинается незамедлительно;
- специалист нажимает кнопку «Start test» - у пользователя появляется кнопка «Start test», при нажатии на которую начинается тестирование.

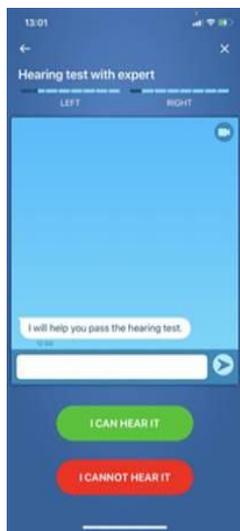
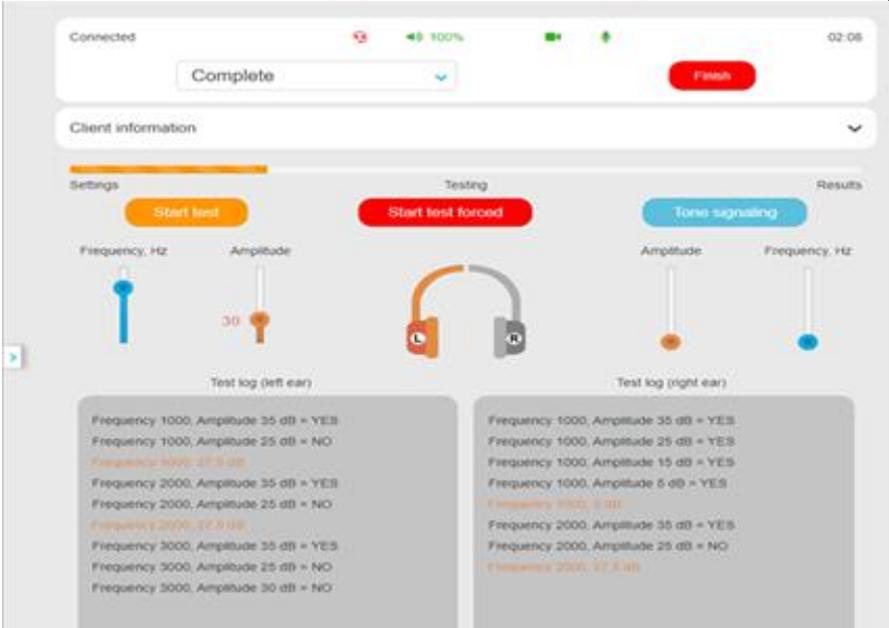
Таблица 4. Графические интерфейсы: Редактирование процесса дистанционного тестирования

Интерфейс пользователя	Интерфейс специалиста
	

Также в интерфейсе пользователя предусмотрены следующие возможности редактирования видео в процессе прохождения дистанционного тестирования слуха (таблица 4):

- кнопка «Обновление видео» - перезапускает видео на стороне пользователя в случае возникновения ошибок;
- кнопки переключения режима звонка чат/видео - переключение происходит с обеих сторон;
- красная кнопка над иконкой показывает выбранный режим у пользователя;
- кнопки отключения микрофона/динамика;
- кнопка отключения трансляции видеосигнала.

Таблица 5. Графические интерфейсы: Интерфейс дистанционного тестирования

Интерфейс пользователя	Интерфейс специалиста
	

При работе в интерфейсе тестирования (таблица 5) специалист видит:

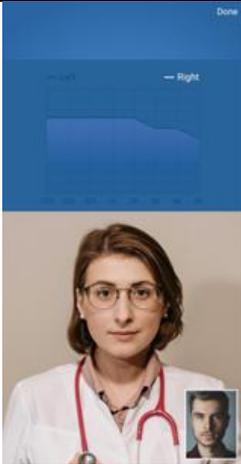
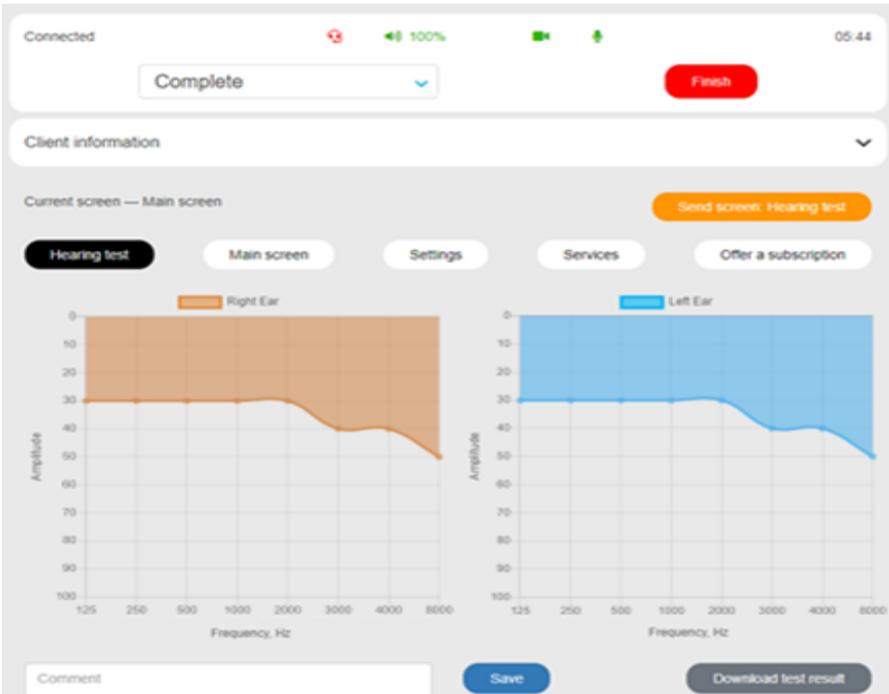
- информацию о прогрессе тестирования, частоте и амплитуде тонального сигнала;

- результаты прошлых тестов.

В интерфейсе «Результаты тестирования слуха» (таблица 6) формируемая автоматически аудиограмма выводится для специалиста и пользователя.

В интерфейсе «Настройка приложения» специалист может демонстрировать для обучения работу возможностей программы, а пользователь - самостоятельно переключать настройки и подстраивать возможности программы индивидуально (таблица 7).

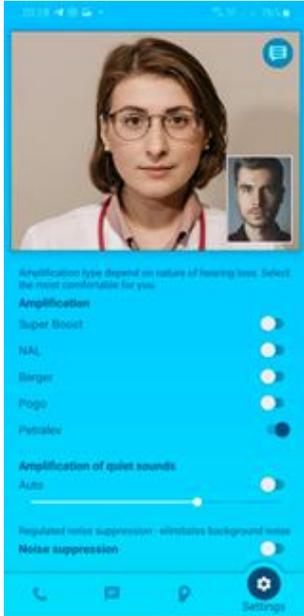
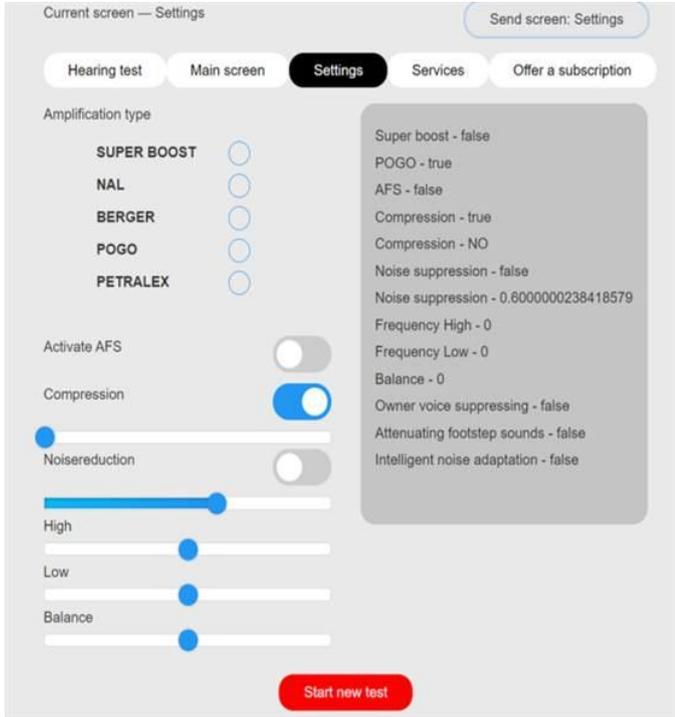
Таблица 6. Графические интерфейсы: Результаты тестирования слуха

Интерфейс пользователя	Интерфейс специалиста
	

После завершения тестирования приложение демонстрирует пользователю и специалисту аудиограмму обеих ушей (таблица 3), которую специалист расшифровывает. После завершения теста, в системе предусмотрена возможность корректировать исходящий от консультанта аудиопоток, с помощью созданного профиля с использованием формулы усиления NAL. Обработка сигнала такого рода имеет целью повысить для пользователя разборчивость речи консультанта в ходе их дальнейшего общения.

Специалист может описать степень потери слуха пользователя с помощью таблицы «звук/расстояние восприятия», которая показана на рисунке 30, где иконками зеленого и красного цвета обозначаются звуки (фоновые шумы: шелест травы, пение птиц, шум воды; шепот на расстоянии 1,5 м, разговорная речь на расстоянии 5 метров). Зеленый цвет означает, что пользователь может слышать звук без затруднений, красный – восприятие звука затруднено.

Таблица 7. Графические интерфейсы: Настройки приложения

Интерфейс пользователя	Интерфейс специалиста
 <p>The screenshot shows a user interface with a video call window at the top. Below it, there is a settings menu for amplification. The menu includes options for 'Amplification type' (Super Boost, NAL, Berger, Pogo, Petralex), 'Amplification of quiet sounds' (Auto), and 'Noise suppression'. A 'Settings' icon is visible at the bottom right.</p>	 <p>The screenshot shows a specialist interface for settings. It includes a top navigation bar with 'Hearing test', 'Main screen', 'Settings', 'Services', and 'Offer a subscription'. The 'Settings' screen is titled 'Current screen — Settings' and 'Send screen: Settings'. It features a list of amplification types (SUPER BOOST, NAL, BERGER, POGO, PETRALEX) with radio buttons. Below this are sliders for 'High', 'Low', and 'Balance', and a 'Start new test' button at the bottom. A detailed list of settings is shown on the right, including 'Super boost - false', 'POGO - true', 'AFS - false', 'Compression - true', 'Compression - NO', 'Noise suppression - false', 'Noise suppression - 0.6000000238418579', 'Frequency High - 0', 'Frequency Low - 0', 'Balance - 0', 'Owner voice suppressing - false', 'Attenuating footstep sounds - false', and 'Intelligent noise adaptation - false'.</p>

После прохождения тестирования специалист может переключаться между основными интерфейсами приложения (главный экран, экран настроек, экран дополнительных сервисов и экран подписки) на устройстве пользователя. Переключив приложение на экран настройки, специалист может продемонстрировать работу разных формул усиления, подавление АОС, компрессию, подавление шума, усиление высоких и низких частот.

Для дистанционного тестирования слуха в приложении используется метод тональной пороговой аудиометрии. Тональная аудиометрия позволяет провести проверку слуха на заданном частотном диапазоне. Результатами оценки слуха является аудиограмма (график, показывающий уровень восприятия звука на разных частотах), которая используется для оценки состояния слуха пациента и для базовой настройки цифровых слуховых аппаратов, в том числе приложений-слуховых аппаратов.



Рисунок 30. Восприятие звуков на расстоянии

Пациенту подается звуковой сигнал определенной частоты и амплитуды, если пациент слышит сигнал, он нажимает кнопку «Слышу» или «Не слышу». После прохождения тестирования специалист расшифровывает аудиограмму и оценивает состояние органов слуха.

При проведении тестирования методом тональной пороговой аудиометрии частота и/или амплитуда сигнала меняется только после ответа пользователя, что позволяет избежать ошибок пользователя, если он неправильно понял инструкцию, и реакции на «фантомные звуки» (состояние, при котором человек полагает, что он слышит звук, который воспроизводит устройство, хотя это не соответствует реальности). В акустических условиях с внешними шумами пороги

слышимости человека на частотах с шумом повышаются – в этом случае проходить тестирование лучше «in situ» («на месте»).

Тестирование методом пороговой тональной аудиометрии начинается при амплитуде 35 дБ с порядком частот 1000, 2000, 4000, 8000, 500, 250, 125 Гц. Если пользователь слышит сигнал и нажимает кнопку «Слышу» - амплитуда понижается на 10 дБ. Далее генерируется сигнал с новой амплитудой и если пользователь не слышит новый сигнал - амплитуда сигнала повышается на 5 дБ.

Если пользователь не слышит сигнал амплитудой 35 дБ - амплитуда повышается до 50 дБ и повышается с шагом 10 дБ, пока не появится реакция или будет достигнут максимальный уровень. Порог слуха определяется при двойном прохождении через минимальный уровень звука, на который реагирует пользователь шагом 10 дБ при понижении уровня и шагом 5 дБ при его повышении.

Таблица 8. Степени нарушения слуха

Частота (Гц)	NAL-R (Целевое усиление, дБ) $X = 0,05$ (HTL500 + HTL1k + HTL2k)	POGO (Целевое усиление, дБ)	Berger (Целевое усиление, дБ)
250	$X + 0,31$ HTL250-17	HTL250/2-10	
500	$X + 0,31$ HTL500-8	HTL500/2-5	HTL500/2,0
750	$X + 0,31$ HTL750-3	HTL750/2-3	HTL750/1,8
1000	$X + 0,31$ HTL1k-1	HTL1k/2	HTL1k/1,6
1500	$X + 0,31$ HTL1,5k-1	HTL1,5k/2	HTL1,5k/1,6
2000	$X + 0,31$ HTL2k-1	HTL2k/2	HTL2k/1,5
3000	$X + 0,31$ HTL3k-2	HTL3k/2	HTL3k/1,7
4000	$X + 0,31$ HTL4k-2	HTL4k/2	HTL4k/1,9
6000	$X + 0,31$ HTL6k-2	HTL6k/2	HTL6k/2,0

Устройство пользователя генерирует тональный сигнал в соответствии со значениями, которое оно получает от Web-интерфейса аудиолога. По окончании тестирования генерируется профиль аудиограммы, на основе которой происходит необходимое усиление, которое рассчитывается на основе,

соответствующих формул усиления (таблица 8), где НТЛ – значения порогов слышимости.

Специалист исходя из данных исследования оценивает степень потери слуха пользователя, которая определяется по лучше слышащему уху. Хотя единая классификация степени тугоухости отсутствует, на практике часто используется таблица 9.

Таблица 9. Степени нарушения слуха

Порог слышимости (дБ)	Степень тугоухости
0-25	Функциональная норма
26-40	Легкая потеря
41-55	Умеренная потеря
56-70	Тяжелая потеря
71-90	Глубокая потеря
>90	Глухота

Выводы по главе 2.

1. Для решения выявленных в первой главе проблем, связанных с влиянием окружающего шума на передачу речи, был разработан метод повышения разборчивости, информативности и натуральности звучания звуковых сигналов, в том числе и музыкальных композиций при прослушивании в акустической шумовой обстановке за счет снижения эффекта маскирования полезного звукового сигнала нестационарными акустическими шумами при помощи использования частотно-зависимого адаптивного усиления.

2. На основе разработанного способа повышения разборчивости и информативности звуковых сигналов в шумовой обстановке был спроектирован и реализован блок обработки с адаптивным компрессором динамического диапазона, которым изменяют динамический диапазон полезного сигнала.

3. Разработан способ компенсации потери слуха в телефонной системе, заключающийся в формировании персонализированных аудиосигналов для слабослышащих пользователей на основе их атрибутов, полученных в результате выполненной аудиограммы, хранящиеся в базе данных на сервере

сотовой сети связи и имеющих привязку к мобильным телефонным номерам слабослышащих пользователей.

4. Разработанный способ отличается от известных тем, что при динамической компрессии формируют комплект субполосных аудиосигналов и управляют динамическим уровнем каждого субполосного аудиосигнала в каждой отдельной неравномерной субполосе частот в соответствии с частотной характеристикой слуха слабослышащего пользователя, коэффициентами алгоритма редактирования шума окружающей среды и функции компрессии динамического диапазона в отдельных неравномерных субполосах.

5. Разработанный способ также имеет отличие в том, что при компенсации акустической обратной связи микшированный аудиосигнал дополнительно микшируют с выходным сигналом модуля компенсации акустической обратной связи, входным сигналом которого служит восстановленный широкополосный аудиосигнал модуля динамической компрессии, при этом расщепляют микшированный аудиосигнал и выходной сигнал модуля динамической компрессии на отдельные частотные каналы, оценивают коэффициенты адаптивной фильтрации в каждом отдельном частотном канале, осуществляют адаптивную фильтрацию, сигнал которой служит выходным сигналом модуля компенсации акустической обратной связи.

6. Разработанный способ позволяет расширить функциональные возможности мобильных мультимедийных устройств (смартфонов и планшетов), мультимедийных воспроизводящих устройств, систем видеоконференций, громкой связи, IP - телефонии, колл-центров, усовершенствовать технические характеристики слуховых аппаратов, корректирующих звуковой сигнал (во время телефонного разговора и при прослушивании аудиоконтента) в агрессивной акустической шумовой обстановке.

7. Разработанный способ позволяет устанавливать в мобильный телефонный аппарат слабослышащего пользователя со встроенным

беспроводным каналом связи разработанные в работе специализированные программные модули – дистанционной настройки приложения устройства компенсации потери слуха с помощью мобильного телефона и приложения дистанционной диагностики слуха.

8. Предложенные и реализованные решения расширяют функционал мобильных устройств и обеспечивают слабослышащим абонентам индивидуальный, качественный и комфортный доступ к ресурсам сети, телекоммуникационных систем, создавая предпосылки для повышения качества жизни людям с ограниченными возможностями с помощью таких современных цифровых технологий как искусственный интеллект, машинное обучение и нейронные сети, в частности, для автоматического распознавания и транскрибирования речи.

Таким образом, разработанные методы и реализованное на их основе программное обеспечение решают существовавшие и описанные в первой главе проблемы клиенто-ориентированного доступа слабослышащих абонентов к ресурсам систем, сетей и устройств телекоммуникации, но не решают проблемы людей с глубокой потерей слуха или глухотой. Обеспечение доступа таких абонентов стала возможной с развитием современных технологий искусственного интеллекта и машинного обучения, одной из которых является технология автоматизированного распознавания речи или технология транскрибирования - преобразования устной речи в текст.

Использование такой технологии для обеспечения доступа к ресурсам систем, сетей и устройств телекоммуникации людей с полной или глубокой потерей слуха является актуальной и перспективной, но не решенной полностью задачей. Связано это и с развитием самой технологии, в которой много нерешенных задач, и со сложностью совмещения такой ресурсозатратной – по вычислительным, временным и информационным показателям с требованиями минимизации ресурсов при использовании в мобильных устройствах.

В следующей главе рассмотрим разработанные новые способы и устройства, предлагающие преодолеть эту проблему.

ГЛАВА 3. РАЗРАБОТКА СИСТЕМЫ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РЕЧИ ДЛЯ МОБИЛЬНЫХ УСТРОЙСТВ

3.1. АНАЛИЗ СУЩЕСТВУЮЩИХ ПРОБЛЕМ ПРИМЕНЕНИЯ СИСТЕМ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ УСТНОЙ РЕЧИ В МОБИЛЬНЫХ УСТРОЙСТВАХ

Современные информационные технологии искусственного интеллекта стремительно завоевывают позиции как в бизнесе – беспилотные автомобили (разработка компании Яндекс), система распознавания лиц, управления сервисами «Digital Manager» (компания Сбер), система управления и контроля непрерывно-травильным агрегатом на Череповецком металлургическом комбинате (компания Северсталь), так и в сферах торговли и обслуживания клиентов – системы взаимодействия с клиентами в том числе и с помощью чат-ботов (компания EORA для Леруа Мерлен), чат-боты-консультанты для онлайн-магазинов, сервисы быстрого поиска по документации и инструменты анализа мнений клиентов (компания iPavlov на базе МИФИ) [83-85].

Особенно эффективно технологии искусственного интеллекта внедряются в государственные структуры – системы управления качеством и автоматизация работы контактных центров, системы биометрического контроля доступа и распознавания лиц (группа компаний ЦРТ), что связано в первую очередь с требованиями большого количества ресурсов для развития и внедрения такой технологии. Искусственный интеллект широко внедряется и в медицину – системы поддержки принятия врачебных решений: платформы обработки электронных медицинских карт, голосовые многофункциональные медицинские терминалы, интеллектуальные системы анализа медицинских снимков (ГК "Maxima") [86].

Под понятие искусственного интеллекта (ИИ) подпадают системы поддержки принятия решения, компьютерное зрение, обработка естественного языка, распознавание и синтез речи. Автоматическое (автоматизированное) распознавание речи (АРР или, в английской терминологии, automatic speech

recognition - ASR) - преобразование звучащей речи в текст, одно из наиболее динамично развивающихся как в теоретических научных исследованиях, так и в практическом применении направление информационных технологий в области искусственного интеллекта и машинного обучения. Разработано множество коммерческих приложений, таких как автоматизация работы call-центров или IVR-системы (Interactive Voice Response) – системы автоматического доступа к информации, минуя оператора для автоматического сбора информации, маркетинговых исследований и других задач [87,88].

Применение систем APP востребовано и в медицинских исследованиях - например, для ввода данных при управлении автономными аппаратами исследования внутренних органов. Важной областью применения систем автоматического распознавания и синтеза речи является помощь людям с инвалидностью, в первую очередь слабовидящим и слабослышающим - для разработки и применения ассистивных технологий [89].

В России медицинские приложения систем автоматического распознавания речи только начинают свое развитие, что делает подобную разработку крайне актуальной. Использование возможностей системы распознавания речи, преобразования речи в текст или автоматический синтез речи при использовании в мобильном телефоне, совмещающем в себе функции слухового аппарата, позволит не только существенно увеличить количественно доступ слабослышающих абонентов к сетям сотовой связи, но и качественным образом изменить сам подход к системам, сетям и устройствам телекоммуникации.

Рассмотрим возможность создания такой системы APP (SAPP) для применения в мобильных устройствах связи, совмещающих функции мобильного телефона и слухового аппарата.

Функционал современных систем APP реализуется в три этапа (рисунок 31) [89]:

1. **Предварительная обработка сигнала.** Система получает голосовой сигнал, например, в формате .wav, записывает и посылает на сервер. Далее происходит определение границ речи, сегментация речевого сигнала, шумоочистка сигнала и деление записи на фонемы. Фонема (от греч. *phōnēma* - звук) - минимальная звуковая единица языка (фрагменты длиной до 25 миллисекунд), линейно не членимая, служащая для образования звуковых оболочек значащих единиц и условно связанная со смыслом звукового строя языка, предельный элемент, выделяемый линейным членением речи. Фонемы в сознании говорящих обычно представлены основными аллофонами. Аллофон (от греч. *allos* - другой звук) – вариант фонемы, обусловленный влиянием на произношение конкретного звука конкретным фонетическим окружением. Замена символов фонемами (аллофонами) осуществляется согласно заданным правилам в соответствии со справочником фонем (аллофонов), который представляет собой хранилище данных: датасет или, чаще всего, базу данных (БД), которые представляют собой наборы аудиозаписей с голосом, которые сопровождаются размеченным текстом. Каждую фонему сервер пропускает через акустическую модель, которая определяет, какие именно звуки (аллофоны) произнесены. Далее выполняется процесс фильтрации фонового шума и анализа звуковых паттернов.

2. **Расшифровка аудио.** Синтагмы записи сравнивают с эталонными произношениями слогов и слов из акустической модели. Синтагма (от греч. *syntagma*, буквально - «вместе построенное, соединенное») - минимальная единица при членении высказывания интонационными средствами, фонетическое целое, выражающее единое смысловое целое в процессе речи-мысли [86]. Синтагма может трактоваться как последовательность аллофонов от паузы до паузы. На этом этапе используются обученные нейронные сети, чтобы подобрать фонетические варианты произнесенных слов и определить их контекст. Для обучения нейронных сетей требуются большие массивы данных (bigdata), хранящих слова в разном произношении, синтагмы, аллофоны,

фонемы. В процессе обучения на вход нейросети подаётся пара аудио-текст, искусственный интеллект разбивает запись с голосом на короткие отрезки и пытается предсказать по спектрограмме каждой из них, что это за синтагма. При этом в процессе предсказания нейросеть не выдаёт однозначный результат: она определяет, с какой вероятностью перед ней та или иная синтагма. Когда вероятности по всем синтагмам записи голоса вычислены, искусственный интеллект с помощью словаря («контекста») определяет слова или фразы, из которых составляет предложения.

3. **Преобразование речи в текст.** С помощью языковой модели алгоритм определяет порядок слов и подбирает нераспознанные слова по контексту. Полученная информация поступает в декодер, который объединяет данные от акустической и языковой моделей и преобразует их в текст. Текст на выходе должен быть связным, осмысленным и правильно оформленным (был поделён на предложения, имел знаки препинания). Связность и осмысленность в технологии распознавания речи обеспечивается, в первую очередь, объёмом текстов, которые нейросеть обработала на этапе обучения.

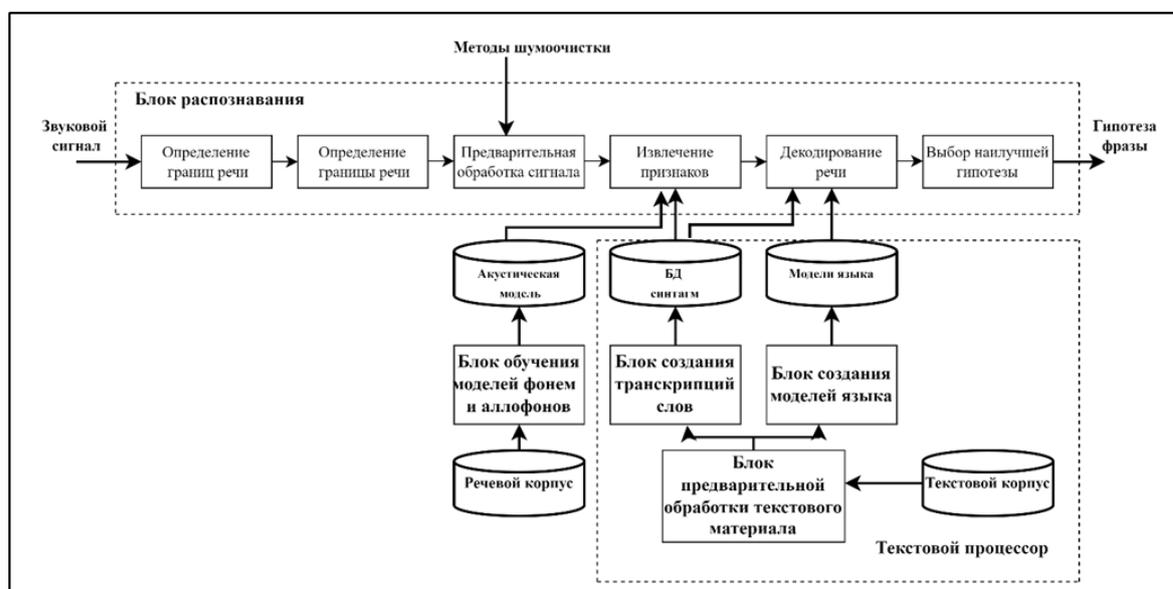


Рисунок 31. Принципиальная структура системы автоматизированного распознавания речи (САРР)

Используемые в САРР методы распознавания базируются на математических, эвристических и статистических методах распознавания речи, предложенными и разработанными как российскими [87-92], так и зарубежными учеными в 70х-90х годах 20 века. Можно выделить две основные технологии, с помощью которых функционируют САРР: традиционный гибридный метод и сквозной метод. Традиционный гибридный метод объединяет подход, основанный на правилах и использующий определенные лингвистические правила, и статистический подход, который опирается на закономерности и взаимосвязи, полученные из больших массивов данных транскрибированного аудио. Несмотря на высокую эффективность, этот гибридный подход может быть очень сложным, но требует сравнительно (относительно нейронных сетей) небольших вычислительных затрат [90-94].

При сквозном методе распознавания речи (end-to-end, E2E) акустическая модель сразу предсказывает текст. САРР обычно используют глубокие нейронные сети (Deep Neural Networks - DNN) для изучения сложных корреляций между аудиосигналом и транскрипцией. После обучения на больших объемах транскрибированного аудио эти системы умело справляются с различными акцентами, произношением и стилем речи. Этот метод устраняет необходимость в явных промежуточных этапах, таких как распознавание фонем или слов, позволяя распознавать речь, учитывая грамматику, синтаксис, структуру и состав фраз, эффективно обрабатывая человеческую речь, что делает метод более эффективным и потенциально точным. Эти системы разработаны таким образом, чтобы учиться на каждом взаимодействии, постоянно совершенствуя свои ответы. Однако технология распознавания фраз в разы увеличивает объемы хранимых обучаемых данных. Для этой цели требуются большие звуковые базы данных (уровня bigdata), содержащие сотни часов записей речи множества дикторов и фонетическую транскрипцию этих записей, которая выполняется автоматически по каноническим правилам и огромные вычислительные ресурсы для обучения.

Помимо этих методов, эффективность и точность САРР определяют многочисленные компоненты и алгоритмы. Взаимодействие этих элементов обеспечивает плавное и точное преобразование речи в текст. Современные САРР вне зависимости от используемого метода способны настраиваться и адаптироваться к конкретным требованиям. Например, повышение точности распознавания можно добиться с помощью языковых весов, подчеркивая часто произносимые слова, такие как профессиональная терминология или наиболее часто встречающиеся слова. Фильтрация «ненормативной», то есть нечасто используемой лексики обеспечивает дополнительный уровень совершенствования, используя фильтры для выявления и очистки определенных слов или фраз в речевом потоке. Также существует возможность - маркировка спикеров, позволяющая при транскрипции приводить или отмечать вклад каждого спикера в беседу с несколькими участниками, что позволит адаптировать разрабатываемую САРР к различным акустическим условиям и стилям речи - к различным тонам, громкости и темпу голоса.

Технологии речевого распознавания нашли свое применение в различных областях. Однако в области поддержки слабослышащих людей множество проблем все еще остаются не решенными, многие идеи требуют дальнейшего развития. Так, программы, работающие с изолированными словами, достигли высокой точности в системах подачи команд голосом – в наиболее распространенных современных приложениях точность распознавания составляет в среднем 95% и зависит в основном от уровня шума. Несмотря на очень значительный прогресс в автоматическом распознавании речи, достигнутом в последние 10 лет, возможности САРР еще очень ограничены по сравнению с человеком. Речь человека обладает высокой степенью изменчивости, связанной как с естественными особенностями произношения, наличием артикуляционных эффектов, приводящих к изменению контекста произносимых фонем, так и с изменениями эмоционального состояния человека, влияющими на спектральный состав и длительность произношения одних и тех

же акустических единиц. Также к изменениям в речевом сигнале приводят условия, в которых находится человек, уровнем шума, помехи различного характера. Наличие речевых сбоев - пауз, хезитаций, повторений, самокоррекций, усеченных слов и т.п. является одним из основных отличий спонтанной речи от письменного текста и сильно влияет на уровень распознавания как звучащей речи, так и ее транскрипций. Перед современными САРР стоит задача распознавания слитной речи. Для этого необходимо реализовать все свойства слухового аппарата человека:

- возможность адаптации и, главное, понимания сказанного вне зависимости от особенностей речи говорящего;
- способность распознавания удаленной, реверберированной, акцентной речи, речи в плохих каналах связи
- способность выделять речь одного диктора из многоголосья и распознавать спонтанную речь.

Однако в реальной речи правила могут нарушаться, а значит, математические структуры, полученные в результате обработки таких баз, не будут описывать речевой сигнал с высокой точностью. Решение этой задачи с достаточно высокими результатами по точности представлено на сегодня только в случае ограниченного словаря. В настоящее время множество работ посвящено проблеме распознавания слитной речи (ИПУ РАН [92] «Истра-Софт» [89], IBM [95], Яндекс [94] и др.), т.к. именно такой тип речевого взаимодействия считается наиболее перспективным и все эти работы связаны с усовершенствованием структуры нейронных сетей, обязательным наличием обратных связей на различных уровнях и разработкой новых методов обучения таких нейронных сетей.

Таким образом, для создания системы автоматического распознавания устной речи, которая могла бы использоваться в мобильных устройствах связи, совмещающих функции мобильного телефона и слухового аппарата необходимо предварительно решить следующие задачи:

1. *Увеличить скорость обработки устной речи.* Для этого необходимо выявить информативные признаки, однозначно характеризующих устную речь, то есть выделить признаки наиболее часто встречающихся в разговорах речевых сигналов для выполнения этапа обработки речи в процессе распознавания. Существует некоторое число математических методов, анализирующих речевой спектр. Здесь самым широко используемым является преобразование Фурье, известное из теории цифровой обработки сигналов. Данный математический аппарат хорошо себя зарекомендовал в данной области, имеется множество методик обработки сигналов, использующих в своей основе преобразование Фурье [97]. Несмотря на это, постоянно ведутся работы по поиску иных путей параметризации речи. Одним из таких новых направлений, является вейвлет-анализ, который стал применяться для исследования речевых сигналов сравнительно недавно. Вейвлет-анализ – это специальный тип линейного преобразования сигналов и представляемых данными сигналами физических данных о процессах и физических свойствах природных средств и объектах. Теория данного метода сейчас развивается учеными всего мира, и многие исследователи возлагают большие надежды на использование инструмента вейвлет-анализа для распознавания речи.

Преимущество вейвлет – анализа перед преобразованием Фурье состоит в том, что оно позволяет проследить за изменением спектральных свойств сигнала со временем и указать, какие частоты (масштабы) доминируют в сигнале. Вейвлеты повсеместно используются с целью фильтрации и предварительной обработки данных, анализа состояния и формирования прогнозов ситуации на фондовых рынках, распознавания образов, при обработке и синтезе разных сигналов, в том числе речевых, для решения проблемы сжатия и обработки изображений, при обучении нейронных сетей и для многих иных целей.

Если рассмотреть речевые распознаватели с позиции классификации по механизму функционирования, то подавляющая их часть относится к системам с вероятностно-сетевыми методами принятия решения о соответствии входного

сигнала эталонному – это метод скрытого Марковского моделирования (СММ), метод динамического программирования и нейросетевой метод. Например, нейронные сети могут быть использованы для классификации характеристик речевого сигнала и принятия решения о принадлежности к той или иной группе эталонов. Нейросеть обладает способностью к статистическому усреднению, т.е. решается проблема с вариативностью речи. Многие нейросетевые алгоритмы осуществляют параллельную обработку информации, т.е. одновременно работают все нейроны. Тем самым решается проблема со скоростью распознавания – обычно время работы нейросети составляет несколько итераций. Сейчас многие разработчики используют аппарат нейронных сетей для построения распознавателей.

2. ***Разработать способ минимизации базы данных фонем, аллофонов и синтагм***, не влияющий на понижение точности обработки текстов. Современные аллофонные базы, используемые в распознавании устной речи или в синтезе речи по тексту, требуют больших объемов памяти и высокой производительности и скорости обработки информации. Такие базы могут содержать мини-набор аллофонов и макси-набор аллофонов [95]. Макси-набор аллофонов более подробный, и при обучении систем синтеза требует большого объема текста. Мини-набор аллофонов менее подробный, но при определенных методиках с большой долей вероятности он позволяет получить всю совокупность аллофонов при чтении диктором меньшего количества фраз из текста.

Известен способ компиляционного фонемного синтеза русской речи и устройство для его реализации [98]. Устройство содержит текстовый процессор, который выполняет следующие функции:

- нормализация текста;
- фонетическая транскрипция по разбивке слова на фонетические единицы по принципу приоритетов;
- идентификация звуковых единиц;

- селекция фонемо-сочетаний вида согласная-гласная-согласная-согласная (СГСС) и согласная-гласная-согласная (СГСконечная);
- организация управления параметрами элементов компиляции и слоговым ударением.

Этот способ реализуется следующим образом. Информация после текстового процессора, освобожденная от цифр и знаков пунктуации, представляет последовательность идентификаторов звуковых единиц, поступающую вместе с признаком ударения на вход акустической базы данных. Одновременно с этим текстовый процессор в результате селекции последовательности типов фонем вида СГСС и СГСконечная вырабатывает признак на формирование фрагмента компиляции СГС, который поступает на блок формирования СГС.

К недостаткам обработки текста по этому способу следует отнести:

- плохое транскрибирование частей слов, т.к. не учитываются соотношения более высокого уровня, следовательно, могут некорректно проставляться словесные ударения, а фразовые просто не проставляются.
- отсутствует информация о паузах, без обработки которых точность обработки текстов снижается.

3.2. ОПИСАНИЕ РАЗРАБОТАННОГО СПОСОБА ФОРМИРОВАНИЯ БАЗЫ ДАННЫХ СБАЛАНСИРОВАННЫХ СИНТАГМ И РАЗРАБОТКА АЛГОРИТМА ВЗАИМОДЕЙСТВИЯ С БАЗОЙ ДАННЫХ

Разработанный способ автоматического распознавания устной речи направлен на распознавание речи с использованием заданной базы фонемных единиц и опирается на способ предварительной обработки текста посредством текстового процессора (рисунок 31), включающий приведение исходного текста в нормализованный орфографический текст путем преобразования сокращений и аббревиатур в линейный текст, членение текста на предложения и слова,

маркировку фразовых и словесных ударений, объединение слов в синтагмы с постановкой символов пауз в конце синтагм с последующим транскрибированием синтагм для получения идеальных транскрипций синтагм в терминах фонем и аллофонов [100]. Границами синтагм являются знаки препинания.

В этом способе предварительной обработки речи к идеальным (каноническим) транскрипциям синтагм (то есть к фонетическим транскрипциям, соответствующим произносительной норме языка) применяют правила транскрипционного моделирования, после применения которых получают дополнительные варианты транскрипций, к которым также применяют правила транскрипционного моделирования. Необходимость транскрипционного моделирования объясняется тем, что произнесение слов и предложений обладает значительной вариативностью. Одно и то же слово, сказанное одним и тем же человеком, может состоять из разных наборов звуков, а в предложениях - паузы могут быть расставлены разным образом.

Причины вариативности произнесения различны. Выделяют такие факторы, как стиль речи, степень формальности речи, наличие акцентов и диалектов, социально-экономические факторы, эмоциональное состояние, анатомические особенности диктора. Из полученного таким образом общего списка исходных и полученных дополнительных вариантов транскрипций исключают одинаковые транскрипции и сохраняют оставшиеся в списке транскрипции для дальнейшего использования.

Ограничение данного способа для использования при распознавании и синтезе речи заключается в том, что в режиме обучения нейронных сетей подбор фраз производится непосредственно диктором, а он не имеет возможности использовать наиболее фонетически соответствующий текст и фразы для представления их своим голосом. Это снижает качество воспроизведения. Кроме того, способ требует для своего осуществления высокой производительности оборудования (скорости обработки информации), поскольку в нем требуется

неоднократное применение достаточно сложных правил транскрипционного моделирования, и в результате получается множество дополнительных вариантов транскрипций, из которых трудно выбрать нужные, фонетически наиболее характерными(сбалансированными) для произносимого текста.

Для преодоления указанных ограничений предлагается известный способ предварительной обработки текста посредством текстового процессора модифицировать следующим образом:

- при получении идеальных транскрипций синтагм в терминах фонем и аллофонов в текстовом процессоре дополнительно формировать базу данных эталонных аллофонов,
- сравнивать совпадение аллофонов идеальных транскрипций синтагм с эталонными аллофонами и исключать аллофоны идеальных транскрипций синтагм, не совпадающих с эталонными аллофонами,
- по аллофонам идеальных транскрипций синтагм, совпадающим с эталонными аллофонами, формировать БД сбалансированных синтагм текста - имеющих наибольшее число совпадений аллофонов идеальных транскрипций синтагм с эталонными аллофонами.

Такое решение, позволяющее сформировать максимально возможное количество вариантов произношения для последующего выбора наиболее близкого к реально произнесенному диктором, возможно реализовать при использовании современных методов транскрипционного моделирования речи.

Транскрипционное моделирование основано на применении правил моделирования речи, список которых формируется как на основании современных знаний о допустимых отклонениях реального произношения от произносительной нормы, так и в результате сбора и обработки статистической информации по разнице реального и нормативного произношении без учета особенностей диктора. Все применяемые правила делятся на две больших группы: правила пропуска звуков, и правила замен звуков. Правила пропуска звуков описывают ситуации, когда звук, который должен присутствовать при

нормативном произнесении высказывания, не произносится. Правила замены звуков описывают ситуации, когда вместо звука, который должен быть произнесен при нормативном произнесении высказывания, произносится какой-то другой звук.

Можно определить правила вставки звуков, описывающие ситуации, в которых при произнесении высказывания в нем появляются дополнительные звуки, отсутствующие в нормативной реализации высказывания. Однако вставка звуков нехарактерна для русского языка, и правила вставки практически не используются. Такой двойной подход к формулированию правил позволяет строить транскрипции, наиболее близкие к произношениям, встречающимся в реальной жизни. Более того, известно, что речь слабослышащих людей отличается своим фонетическим звучанием – такой подход позволит создать базу данных специализированных транскрипций («сбалансированных синтагм»).

Также целесообразно, чтобы:

- сбалансированные синтагмы формировались в виде таблицы в порядке их сбалансированности;
- ограничивалось количество сбалансированных синтагм;
- задавался минимальный процент количества сбалансированных синтагм к общему количеству синтагм;
- производился бы процесс уменьшения базы данных эталонных аллофонов, в котором для наиболее сбалансированной синтагмы текста из базы данных эталонных аллофонов исключают эталонные аллофоны, содержащиеся в наиболее сбалансированной синтагме текста, затем для формируемой следующей по сбалансированности синтагмы текста из базы данных эталонных аллофонов исключают эталонные аллофоны, содержащиеся в ней, процесс уменьшения базы данных эталонных аллофонов повторяют для последующих по сбалансированности синтагм, достигая заданного количества сбалансированных

синтагм или заданного процента количества сбалансированных синтагм к общему количеству синтагм.

Реализация этих требований позволит уменьшить вычислительные ресурсы, необходимые для автоматизированного транскрибирования устной речи, что, в свою очередь, позволит использовать такие возможности в мобильных устройствах.

Блок-схема алгоритма формирования сбалансированных синтагм представлена на рисунок 32.



Рисунок 32. Часть блок-схемы алгоритма формирования сбалансированных синтагм

Исходными данными являются идеальные транскрипции синтагм в терминах фонем и аллофонов по совпадающим аллофонам, а также список «совпадающие аллофоны: их количество».

Блок 21 принимает эти данные, а также данные из БД о заданных параметрах минимального % сбалансированности синтагм или минимального количества сбалансированных синтагм на его управляющие входы и осуществляет поиск синтагм, имеющих наибольшее количество совпадающих аллофонов (исходного текста и эталонных). Если не достигнуто заданное количество синтагм или % сбалансированности, то с выхода «Нет» блока принятия решения данные об этом поступают на третий управляющий вход блока 21, а блок 21 осуществляет поиск следующей синтагмы по совпадающим

аллофонам. Если достигнуто заданное количество синтагм или минимальный % сбалансированности, то с выхода «Да» блока принятия решения данные поступают в блок 22, который формирует список сбалансированных синтагм. Таким образом, сбалансированные синтагмы можно формировать в виде таблицы в порядке их сбалансированности, и/или задавать общее количество сбалансированных синтагм, и/или задавать процент количества сбалансированных синтагм к общему количеству синтагм.

Кроме того, для уменьшения баз данных эталонных аллофонов и ускорения процесса формирования сбалансированных синтагм можно для наиболее сбалансированной синтагмы текста из базы данных эталонных аллофонов исключить эталонные аллофоны, содержащиеся в наиболее сбалансированной синтагме текста в БД. Для следующей по сбалансированности синтагмы текста из базы данных эталонных аллофонов исключают эталонные аллофоны, содержащиеся в ней. Процесс уменьшения базы данных эталонных аллофонов повторяют для последующих по сбалансированности синтагм, достигая заданного количества сбалансированных синтагм или заданного процента количества сбалансированных синтагм к общему количеству синтагм.

Далее, после уже после выявленных сбалансированных синтагм для другого фрагмента текста заявленный способ можно повторить. По сформированной в текстовом процессоре уменьшенной базе данных эталонных аллофонов сравнивают совпадение аллофонов идеальных транскрипций синтагм с эталонными аллофонами и исключают фонемы и аллофоны идеальных транскрипций синтагм, несовпадающие с эталонными аллофонами. По аллофонам идеальных транскрипций синтагм, совпадающим с эталонными аллофонами, формируют сбалансированные синтагмы текста - имеющие наибольшее число совпадений аллофонов идеальных транскрипций синтагм с эталонными аллофонами.

3.3. РАЗРАБОТКА СИСТЕМЫ ТРАНСКРИБИРОВАНИЯ УСТНОЙ РЕЧИ С ИСПОЛЬЗОВАНИЕМ БАЗЫ ДАННЫХ СБАЛАНСИРОВАННЫХ СИНТАГМ

Разработанный способ распознавания речи с использованием заданной базы фонемных единиц опирается на известный способ предварительной обработки текста посредством текстового процессора, лежащий в основе функционирования компьютерного устройства для обработки текста [101]. Это устройство предназначено для использования слепыми и как средство обучения русскому языку. Оно позволяет обеспечить высокое качество распознавания и синтеза русской речи при воспроизведении плоскопечатных текстов.

Устройство имеет блок ввода текста, который выполнен оптически, для распознавания плоскопечатного текста, блок анализа, входящий в состав блока синтеза русской речи по орфографическому тексту, блок базы данных и блок представления результатов, выполненный в виде тактильного дисплея. Кроме того, устройство содержит блок формирования аудиосигнала, блок унификации текстового файла, блок сопряжения тактильного дисплея с персональным компьютером и блок интерфейса. Все недостатки способа предварительной обработки текста посредством текстового процессора, указанные выше, присущи и рассматриваемому устройству.

Для устранения этих недостатков разработанное компьютерное устройство для обработки текста (рисунок 33) имеет следующую структуру:

- введён блок ввода параметров (блок 5),
- введён блок формирования сбалансированных синтагм (блок 6),
- выход блока ввода параметров соединен с входом блока базы данных (блок 3), предназначенным для формирования базы данных эталонных аллофонов,
- выход блока анализа (блок 2) соединен с вторым входом блока базы данных,

- выход блока базы данных соединен с входом блока формирования сбалансированных синтагм - такими, которые имеют наибольшее число совпадений аллофонов текста с эталонными аллофонами, выход которого соединен с входом блока представления результатов.

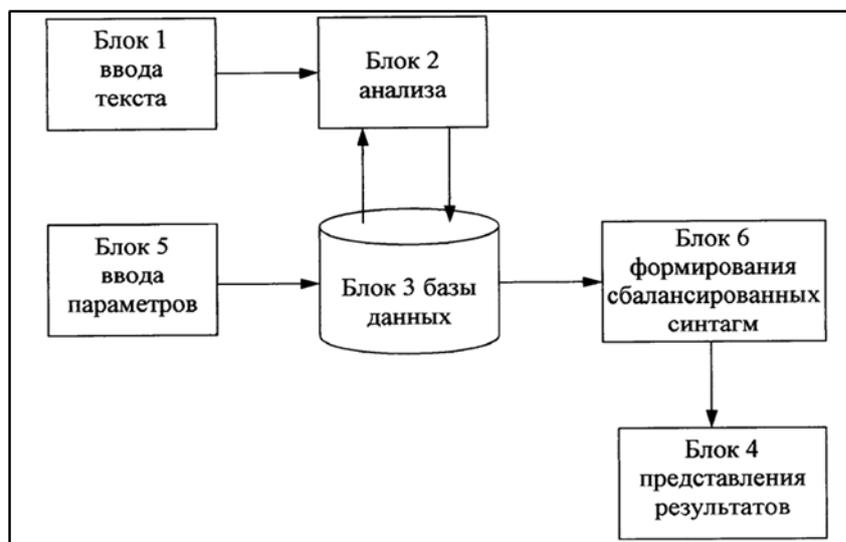


Рисунок 33. Функциональная схема устройства

Блок 1 ввода текста служит для загрузки анализируемого текста из текстового файла или с помощью устройств его ввода (клавиатура, сканер и т.п.).

Блок 2 анализа предназначен для:

- формирования на основе анализируемого текста синтагм;
- замены (отображения) символов (букв) синтагм на фонемы;
- замены (отображения) фонем на аллофоны;
- поиск в тексте аллофонов, совпадающих с эталонными аллофонами;
- определения количества совпадающих аллофонов в анализируемом тексте (то есть определения набора записей вида: «совпадающий аллофон текста с эталонным» - «их количество в тексте»).

Блок 3 базы данных служит для хранения следующей информации:

- параметров анализа текста;
- словаря ударений;
- списка эталонных аллофонов;

- списка совпадающих аллофонов - их количество в тексте;
- результатов анализа текста по совпадающим аллофонам.

Блок 4 представления результатов предназначен для представления пользователю результатов автоматизированного фонетического анализа текста. Результатом анализа текста является выделенный из него набор наиболее фонетически сбалансированных синтагм. Отображение результатов анализа текста пользователю может осуществляться с помощью различных устройств вывода информации (монитор, принтер и т.п.).

Блок 5 ввода параметров служит для ввода пользователем параметров анализа текста с помощью устройства ввода (клавиатуры, «мыши» и т.п.). Параметрами анализа текста являются:

- количество выводимых в результатах поиска сбалансированных синтагм,
- минимальный суммарный процент сбалансированности синтагм,
- алгоритм анализа текста (соответствующее программное обеспечение).

Блок 6 формирования сбалансированных синтагм предназначен для создания по совпадающим аллофонам сбалансированных синтагм - [фраз (предложений)], имеющих наибольшее число совпадений аллофонов текста с эталонными аллофонами блока 3.

Функционирование реализованного устройства в соответствии с разработанным способом распознавания или синтеза речи (рисунок 32), организовано следующим образом (блок-схема алгоритма приведена на рисунке 34):

1. Анализируемый текст поступает с блока 1 на первый вход блока 2 анализа. С блока 5 ввода параметров в блок 3 базы данных поступают параметры анализа текста, список эталонных аллофонов, словарь ударений, которые сохраняются в блоке 3, и далее поступают на второй вход блока 2 анализа.

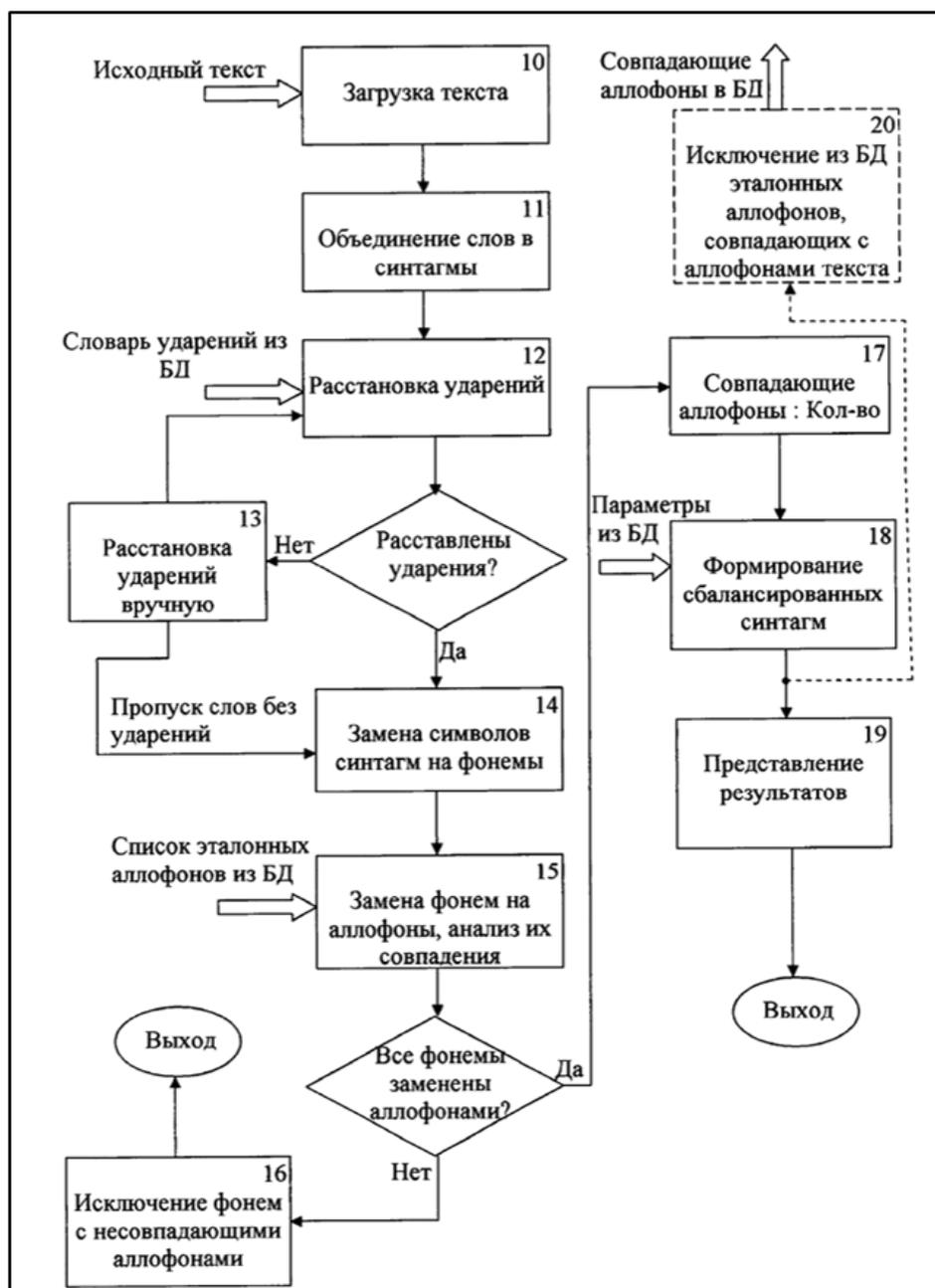


Рисунок 34. Блок-схема алгоритма работы устройства

2. Блок 2 осуществляет приведение исходного текста в нормализованный орфографический текст путем преобразования сокращений и аббревиатур в линейный текст. Затем блок 2 производит членение текста на предложения и слова, маркировку фразовых и словесных ударений, объединение слов в синтагмы с простановкой символов пауз вконец синтагм. После формирования синтагм блоком 2 производится их транскрибирование для получения идеальных транскрипций синтагм в терминах фонем и аллофонов.

Блок 2 сравнивает совпадение аллофонов исходного текста с эталонными аллофонами, исключает аллофоны текста, не совпадающие с эталонными аллофонами. Затем блок 2 составляет список: совпадающие аллофоны - их количество, который поступает в блок 3.

3. Из блока 3 этот список поступает в блок 6 формирования сбалансированных синтагм, который по существу производит обратное преобразование текста относительно операции транскрибирования, осуществляемой блоком 3: из аллофонов формируются фонемы, затем определяются сбалансированные синтагмы - имеющие наибольшее число совпадений аллофонов исходного текста с эталонными аллофонами.

4. На выходе блока 6 по совпадающим аллофонам формируют список фонетически сбалансированных синтагм в зависимости от количества совпадений аллофонов. Под эталонными аллофонами блока 3 в настоящем изобретении понимаются базы данных аллофонов, сформированные в соответствии с методом создания мини-набора аллофонов или макси-набора аллофонов.

5. Блок 10 осуществляет загрузку исходного текста, приведение исходного текста в нормализованный орфографический текст путем преобразования сокращений и аббревиатур в линейный текст. Затем блок 10 производит членение текста на предложения и слова.

6. Блок 11 производит анализ линейного текста и объединение слов в синтагмы. Синтагмы поступают в блок 12, который производит маркировку фразовых и словесных ударений. Расстановка ударений в символах синтагмы производится в соответствии со словарем ударений, поступающим из базы данных (БД) блока 3, куда он вводится с помощью блока 5.

7. Блок принятия решения «Расставлены ударения?» осуществляет проверку выполнения ударений, и, если это нужно, то с его выхода «нет» (ударения не расставлены) предложение о расстановке ударений поступает в блок 13, один из выходов которого служит для пропуска слов без ударений,

связанным с входом блока 14 замены символов синтагм на фонемы. Другой выход блока 13 подсоединен к второму входу блока 12, и ударения могут быть расставлены вручную.

8. С выхода «Да» блока принятия решения «Расставлены ударения?» данные синтагм поступают на блок 14 замены символов (букв) синтагм на фонемы.

9. Далее фонемы в синтагмах поступают на блок 15 замены фонем на аллофоны в соответствии со списком эталонных аллофонов, поступающих из БД блока 3, куда они вводятся с помощью блока 5. На выходе блока 15 имеем идеальные транскрипции синтагм в терминах фонем и аллофонов.

Блоком принятия решения «Все фонемы заменены аллофонами?» производится сортировка синтагм по совпадающим аллофонам. Если аллофоны синтагм исходного текста не имеют совпадений с эталонными аллофонами, то с выхода «Нет» этого блока принятия решения данные поступают в блок 16 исключения фонем с несовпадающими аллофонами. Если все аллофоны синтагм исходного текста имеют совпадения с эталонными аллофонами, то с выхода этого блока принятия решения данные о совпадающих аллофонах поступают на вход блока 17.

10. Блок 17 формирует список «совпадающие аллофоны: их количество». Этот список поступает в блок 18 формирования сбалансированных синтагм, на который также поступают из БД параметры:

- исходный текст,
- минимальный % сбалансированности синтагм или их количество.

11. Блок 18 осуществляет поиск синтагм с наибольшим количеством совпадающих аллофонов, которые соответственно поступают на блок 19 представления результатов и на выход. Для уменьшения («сужения») БД эталонных аллофонов с выхода блока 18 формирования сбалансированных синтагм список аллофонов может быть дополнительно подан в блок 20 исключения из БД эталонных аллофонов, совпадающих с аллофонами текста для

уменьшения базы данных эталонных аллофонов. Этим достигается дополнительное уменьшение информационных ресурсов и ускорение процесса обработки информационных данных. Этот список соответственно передается в БД.

Пользователь запускает специальное программное обеспечение на компьютерном устройстве для обработки текста (рисунок 32). Отображается графический интерфейс (рисунок 35) в виде диалогового окна, которое содержит инструменты (кнопки) 30, 31, 32, 33, 34.

Инструмент 30 «Выделение аллофонов» служит для загрузки исходного текста из текстового файла на диске, инструмент 31 «Настройки» - для редактирования словаря ударений и списка эталонных аллофонов, инструмент 32 «Текст» - для отображения области представления результатов выделения из текста синтагм, фонем и аллофонов, инструмент 33 «Аллофоны» - для отображения области представления таблицы вида «Фонема - Аллофоны - Найденное количество в тексте», инструмент 34 «График» - для визуального графического анализа сбалансированности текста.

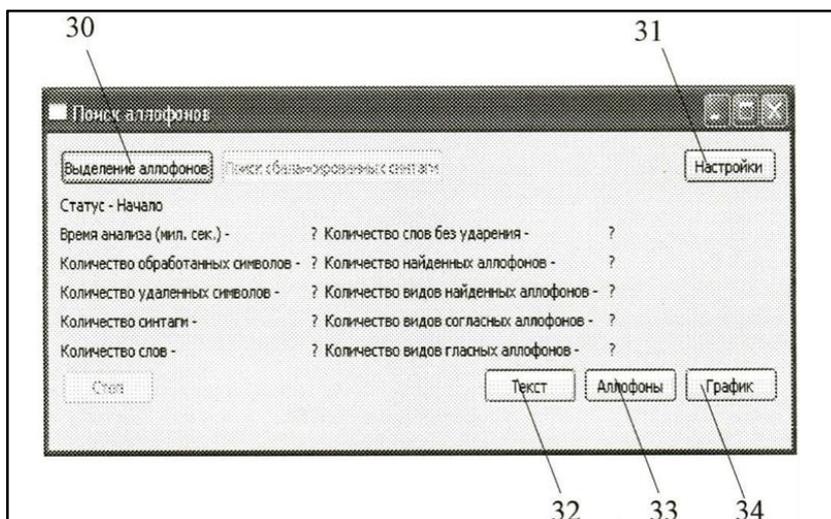


Рисунок 35. Графический интерфейс для ввода текстового файла

Для загрузки текста из файла пользователь нажимает кнопку «Выделение аллофонов». В отобразившемся окне (рисунок 36) пользователь инструментом 35 указывает: язык анализируемого текста в выпадающем списке «Выберите

язык», инструментом 36 полный путь к файлу исходного текста в поле данных «Укажите файл». Поле данных списка аллофонов инструмента 37 служит для выбора мини-набора эталонных аллофонов и макси-набора эталонных аллофонов. Инструмент 38 «Начать» служит для применения сделанных установок.

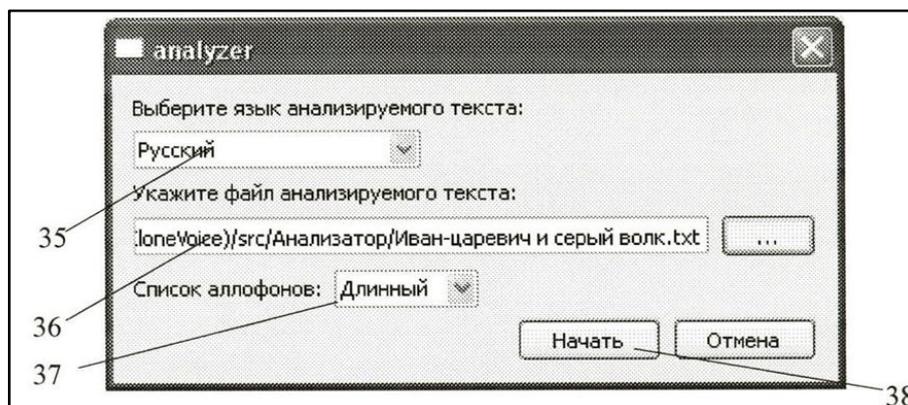


Рисунок 36. Графический интерфейс для указания языка текста и пути к его файлу

Блок 2 анализа (рисунок 32) текста осуществляет разбор введенного пользователем на предыдущем шаге исходного текста на синтагмы. Расставляются ударения в словах, входящих в состав каждой выделенной синтагмы. Расстановка ударений осуществляется с помощью словаря ударений, содержащегося в блоке 3 базы данных.

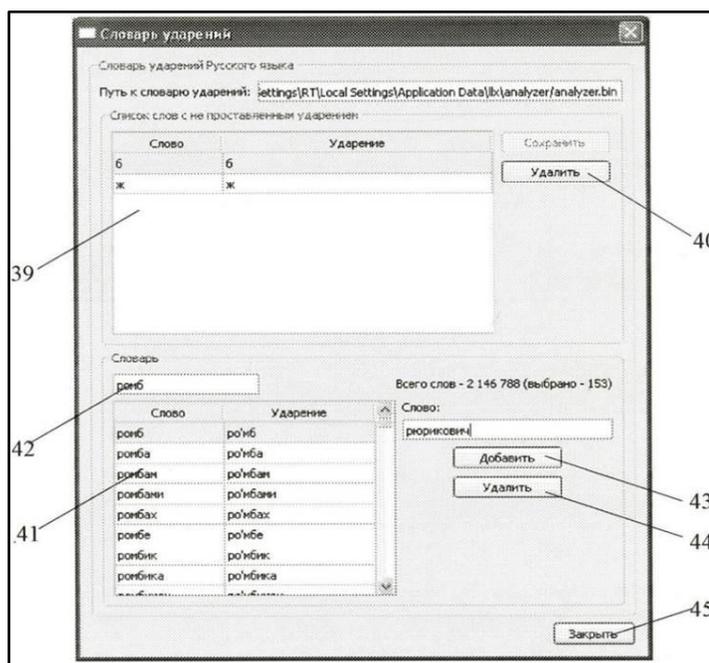


Рисунок 37. Графический интерфейс для редактирования словаря ударений

Также словарь ударений может быть отредактирован пользователем. Для редактирования словаря ударений пользователь в отображенном графическом интерфейсе (рисунок 37) нажимает инструмент 31 «Настройки» (переход)→«Словарь ударений». В отобразившемся графическом интерфейсе (рисунок 37) пользователь осуществляет редактирование словаря ударений. Поле данных 39 служит для составления списка слов с непроставленными ударениями. Инструмент 40 «Удалить» предназначен для удаления слова из поля данных 39 для дальнейшего проставления ударений вручную. Поле данных 41 служит для простановки ударений вручную и составления списка слов для слова, введенного в поле данных 42. Инструменты 43, 44 служат для добавления слова или его удаления соответственно.

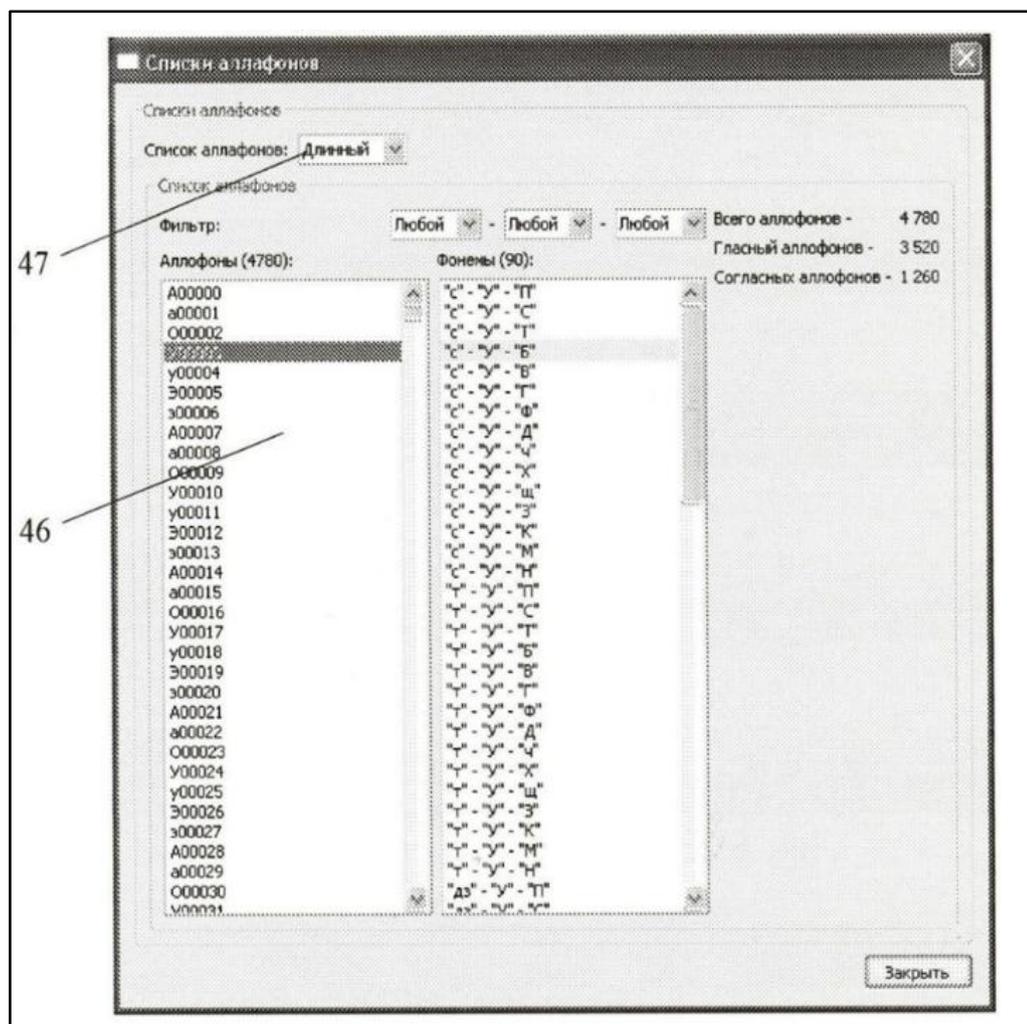


Рисунок 38. Графический интерфейс для редактирования списка аллофонов

Инструмент 45 «Закрывать» предназначен для введения словаря расставленных вручную ударений, в блок 3 базы данных с помощью блока 5 (рисунок 32). Блок 2 анализа текста осуществляет замену символов (букв) синтагмы на фонемы, а фонем на аллофоны. Замена фонем исходного текста на аллофоны осуществляется согласно списку эталонных аллофонов, содержащемуся в блоке 2 базы данных.

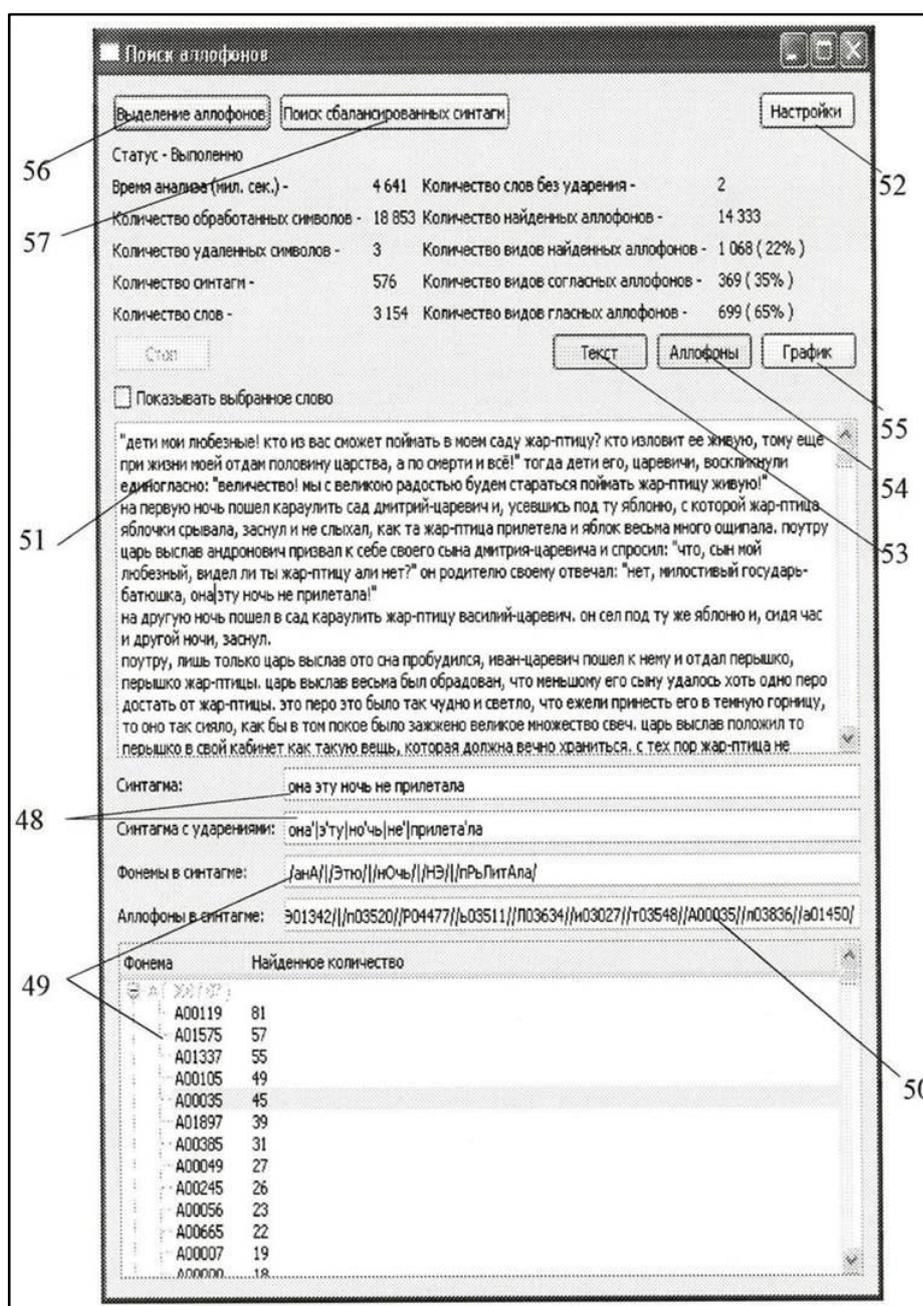


Рисунок 39. Графический интерфейс для поиска сбалансированных синтагм

Список аллофонов может быть также отредактирован пользователем. Для редактирования списка эталонных аллофонов пользователь в графическом интерфейсе (рисунок 35) нажимает инструмент 31 «Настройки» (переход)→«Списки аллофонов». В отобразившемся графическом интерфейсе (рисунок 38) в поле данных 46 пользователь осуществляет редактирование списка аллофонов при выборе мини- набора эталонных аллофонов или макси- набора эталонных аллофонов в поле данных 47.

Результат выделения аллофонов из анализируемого текста пользователь может просмотреть, нажав инструменты 32, 33 «Текст» и «Аллофоны» (рисунок 35). Результат выделения синтагм, фонем, аллофонов отображается в графическом интерфейсе (рисунок 39) в полях данных 48, 49, 50 соответственно. В поле данных 51 отображается исходный текст. Инструмент 52 «Настройки» служит для редактирования словаря ударений и списка эталонных аллофонов, инструмент 54 «Аллофоны» - для отображения области представления результатов выделения из текста синтагм, фонем и аллофонов, инструмент 55 «График» - для визуального графического анализа сбалансированности текста, инструмент 56 «Выделение аллофонов» - для загрузки исходного текста из текстового файла на диске.

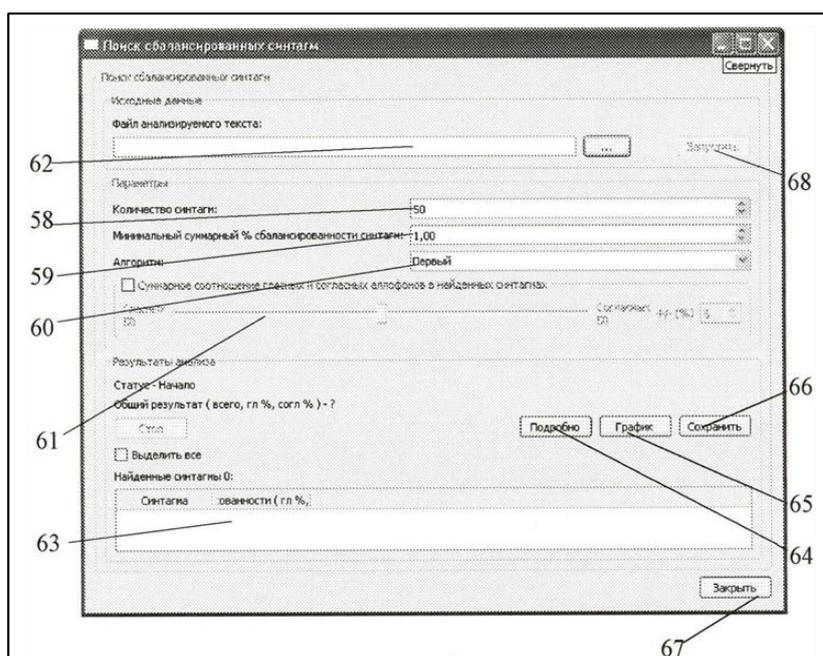


Рисунок 40. Графический интерфейс для ввода параметров анализа текста

Для поиска сбалансированных синтагм пользователь в отображенном графическом интерфейсе нажимает инструмент 57 (кнопку) «Поиск сбалансированных синтагм». Отображается графический интерфейс, вид которого представлен на рисунке 40. В отображенном графическом интерфейсе пользователь указывает следующие параметры анализа текста:

- в поле данных 58 - количество синтагм (с наилучшим фонетическим балансом);
- в поле данных 59 - минимальный суммарный процент сбалансированности синтагм;
- в поле данных 60 - алгоритм анализа текста, (первый или второй, подробнее об алгоритмах смотри ниже);
- в поле данных 61 - суммарное соотношение гласных и согласных аллофонов в найденных синтагмах;
- в поле данных 62 - поле ввода пути и имени файла с исходным текстом;
- в поле данных 63 - таблица вида «Синтагма»- «% согласованности (% гласных, % согласных)».

Инструмент 64 «Подробно» служит для отображения блока «Подробно о синтагме», содержащего: поле «Синтагма», поле «Синтагма с ударениями», поле «Фонемы в синтагме», поле «Аллофоны в синтагме», список «Совпавших аллофонов», список «Не совпавших аллофонов».

Инструмент 65 «График» предназначен для графического представления результатов анализа, инструмент 66 «Сохранить» - для сохранения результата анализа (таблицы сбалансированных синтагм вида «Синтагма» - «% согласованности (% гласных, % согласных)») в текстовом файле на диске, инструмент 67 «Закрыть» - для закрытия окна «Поиск сбалансированных синтагм».

Если в отображенном графическом интерфейсе (рисунок 40) пользователь нажимает инструмент 68 «Запустить», то данный набор сбалансированных

синтагм отображается пользователю на экране монитора компьютерного устройства (рисунок 41) в поле данных 69.

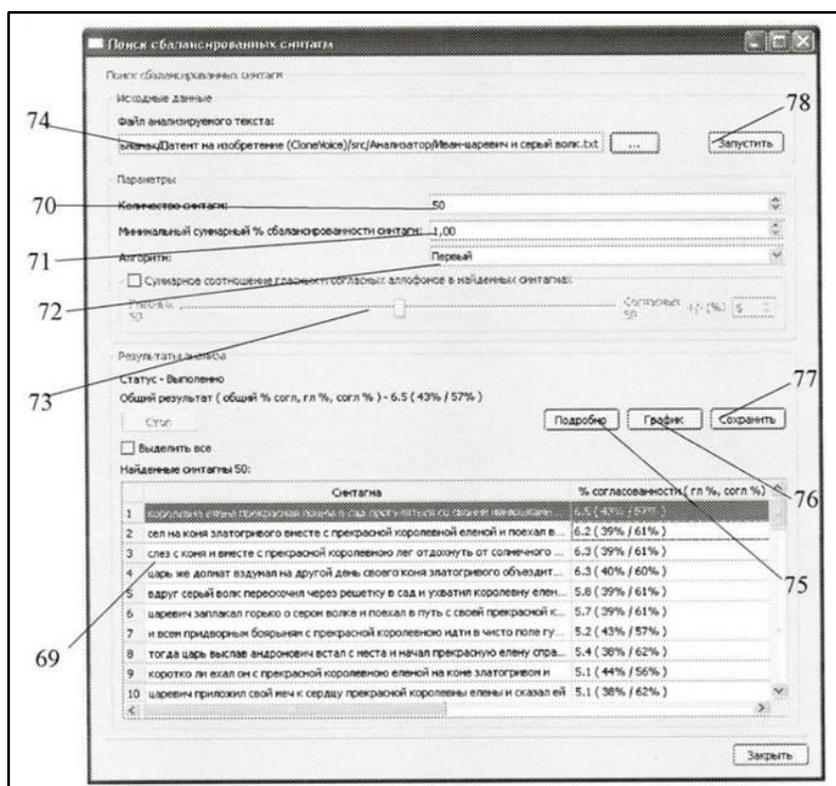


Рисунок 41. Графический интерфейс сформированных сбалансированных синтагм

Функции полей данных 70÷74 и инструментов 75÷78 соответствуют полям данных и инструментам, показанным на рисунке 41. Блок 2 анализа текста (рисунок 32) осуществляет определение количества найденных в исходном аллофонов, совпадающих с эталонными аллофонами, их уникальность и частоту появления в тексте. Результатом данного анализа является составленный и сохраненный в базе данных блока 3 список вида: «совпадающие аллофоны» - «их количество в тексте». Блок 6 поиска сбалансированных синтагм осуществляет анализ и выборку из анализируемого текста синтагм, которые являются наиболее сбалансированными и наилучшим образом фонетически характеризуют исходный текст.

Анализ текста может быть выполнен различными способами:

1. Выделение из текста синтагм с наилучшим фонетическим балансом (то есть содержащие наибольшее количество совпадающих аллофонов) в порядке их сбалансированности. Количество таких синтагм ограничено пользовательской настройкой (количество синтагм) или системой в зависимости от заданного пользователем процента сбалансированности синтагм (минимальный суммарный процент сбалансированности синтагм). Первый алгоритм позволяет получить наилучшее качество воспроизведения диктором исходного текста, но требует большего времени на обработку данных.

2. Анализ процента покрытия базы эталонных аллофонов системы аллофонами, найденными в тексте (отношение количества аллофонов в тексте к количеству эталонных аллофонов в базе системы). Те аллофоны из базы системы, которые отсутствуют в тексте, не учитываются при дальнейшем анализе (база рассматриваемых аллофонов «сужается»). Определяется самая сбалансированная синтагма в тексте (содержащая самый высокий процент эталонных аллофонов из базы). Из базы эталонных аллофонов исключаются аллофоны, содержащиеся в выявленной наиболее сбалансированной синтагме.

Далее в тексте определяется следующая по сбалансированности сбалансированная синтагма и аналогичным образом выполняется «сужение» базы эталонных аллофонов. Процесс «сужения» базы эталонных аллофонов повторяется до тех пор, пока не будет достигнуто заданное количество синтагм или минимальный суммарный процент сбалансированности синтагм. Второй алгоритм позволяет сократить время на обработку математически формализованных данных текста.

Разработанное компьютерное устройство позволяет использовать для анализа текста и другие алгоритмы.

Выводы по главе 3.

1. Был проведен анализ структуры и функциональных блоков современных систем автоматизированного распознавания устной речи, который

позволил сделать вывод о возможности использования такой системы в мобильных телефонах с функцией слухового аппарата.

2. Были выявлены задачи, реализация которых позволит уменьшить объем информационных ресурсов для САРР в мобильных телефонах с функцией слухового аппарата речи:

- увеличить скорость обработки устной речи;
- разработать способ минимизации БД фонем, аллофонов и синтагм.

3. Разработан способ формирования сбалансированных синтагм, имеющих наибольшее число совпадений аллофонов идеальных транскрипций синтагм с эталонными аллофонами для предварительной обработки текста системой автоматизированного распознавания текста.

4. На основе разработанного способа было спроектировано и реализовано техническое и программное обеспечение устройства для транскрипции устной речи, ориентированное на использование в мобильных устройствах с функцией слухового аппарата при передачи речевых сигналов по сети сотовой связи.

5. Разработанное устройство позволяет распознавать наиболее часто используемые синтагмы устной речи с уменьшением вычислительных и информационных ресурсов в среднем на 25% за счет уменьшения объемов БД.

6. Использование такого устройства позволит абонентам с полным отсутствием слуха иметь доступ к ресурсам систем, сетей телекоммуникации, а слабослышащим абонентам – дублировать речь собеседника в виде текста, увеличивая таким образом уровень разборчивости речевых сигналов. й.

7. Применение разработанного устройства позволит расширить функциональные возможности мобильных устройств не только в качестве ассистивной технологии поддержки слабослышащих и полностью лишенных слуха абонентов, но и всем абонентам систем, сетей и устройств телекоммуникации, позволяя на основе предложенного программно-аппаратного комплекса транскрибирования устной речи разрабатывать и

внедрять современные инновационные клиенто-ориентированные мобильные приложения.

Таким образом, был разработан способ уменьшения вычислительных и информационных ресурсов для применения САРР в мобильных устройствах слабослышащих абонентов за счет фильтрации хранимых данных в БД САРР. Необходимо определить, влияет ли такое уменьшение вычислительных и информационных ресурсов на точность распознавания устной речи. Исследуем эффективность этого и предложенных ранее решений в следующей главе.

ГЛАВА 4. ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ПРЕДЛОЖЕННЫХ ТЕХНИЧЕСКИХ РЕШЕНИЙ

4.1 ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ РАЗРАБОТАННОГО СПОСОБА ПОВЫШЕНИЯ РАЗБОРЧИВОСТИ ПОЛЕЗНОГО СИГНАЛА В АГРЕССИВНОЙ ШУМОВОЙ ОБСТАНОВКЕ

Исследование проводилось с помощью экспериментального метода [104-106]. Был разработан и проведен эксперимент, позволяющий оценить повышение разборчивости полезного сигнала $x[n]$ в агрессивной шумовой обстановке. Для проведения эксперимента были заданы следующие условия:

- полезный сигнал $x[n]$ является речевым (его спектограмма представлена на рисунке 42-а). В качестве полезного сигнала $x[n]$ может выступать сигнал мультимедийных воспроизводящих устройств, сигнал в системах видеоконференций, громкой связи, IP-телефонии и т.п. В эксперименте полезный сигнал $x[n]$ - сигнал дальнего диктора, приходящий из сети мобильной связи, а «Слушатель» - ближний диктор, для которого восприятие сигнала дальнего диктора $x[n]$ затруднено акустическим шумом $v[n]$ окружающей обстановки, в которой он находится;

- шумовой сигнал $v[n]$ акустической обстановки - шум метро (спектограмма – рисунок 42-б);

- частота дискретизации была задана 44,1 кГц.

На рисунке 42-б проиллюстрирован полученный эффект предложенного способа. Как видно из рисунка 42-а, полезный сигнал $x[n]$ на выходе блока обработки - выходной сигнал $y[n]$ не имеет никаких артефактов при отсутствии шумового сигнала $v[n]$. Анализ спектрограммы (рисунок 42-б) сигнала на выходе громкоговорителя, который представляет собой сумму выходного сигнала $y[n]$ и шума $v[n]$, и который воспринимает слушатель, показывает, что полезный

сигнал $x[n]$ (речевой сигнал) можно идентифицировать после обработки согласно предлагаемому способу.

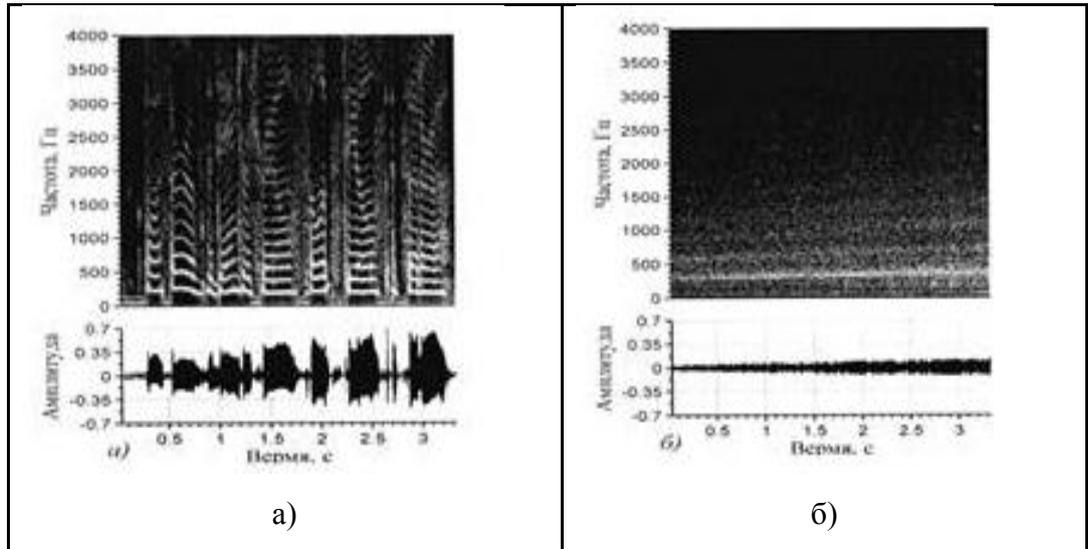


Рисунок 42. Частотно-временное представление сигнала (спектрограммы): а) чистый речевой сигнал; б) сигнал шума метрополитена

Для сравнения на рисунке 43-а представлена спектрограмма полученного до применения предложенного способа сигнала, когда при покрытии чистого речевого сигнала шумом метрополитена этот речевой сигнал едва может быть идентифицирован.

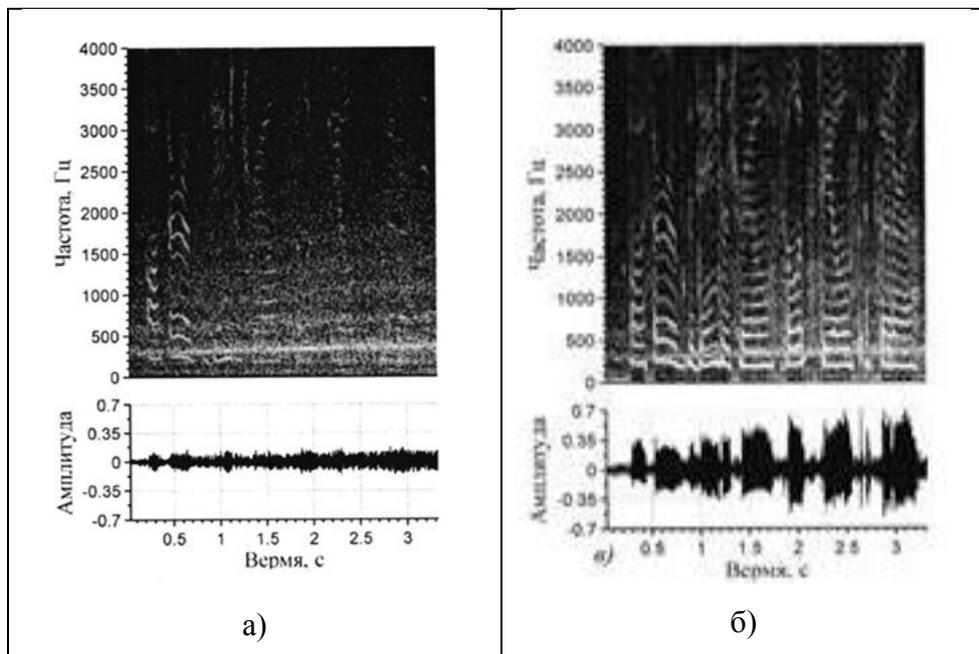


Рисунок 43. Частотно-временное представление сигнала (спектрограммы), который воспринимает слушатель: сумма выходного сигнала $y[n]$ и шума $v[n]$: а) без обработки предложенным способом; б) после обработки способом повышения разборчивости

Таким образом, полезный сигнал $x[n]$ продолжает быть слышимым даже при высокой интенсивности шумового сигнала $v[n]$ акустической обстановки.

Предложенный способ повышения разборчивости и информативности звуковых сигналов в шумовой обстановке экспериментально проверялся при различных отношениях сигнал/шум (ОСШ), во всех случаях измерялась разборчивость речевого сигнала (измерялся индекс разборчивости СИ) на фоне шумового сигнала $v[n]$ для метрополитена разной интенсивности до обработки, когда выходной сигнал $y[n]$ равнялся полезному сигналу $x[n]$ и после обработки банками фильтров и АКДД. Полученные результаты экспериментов изображены на рисунке 44.

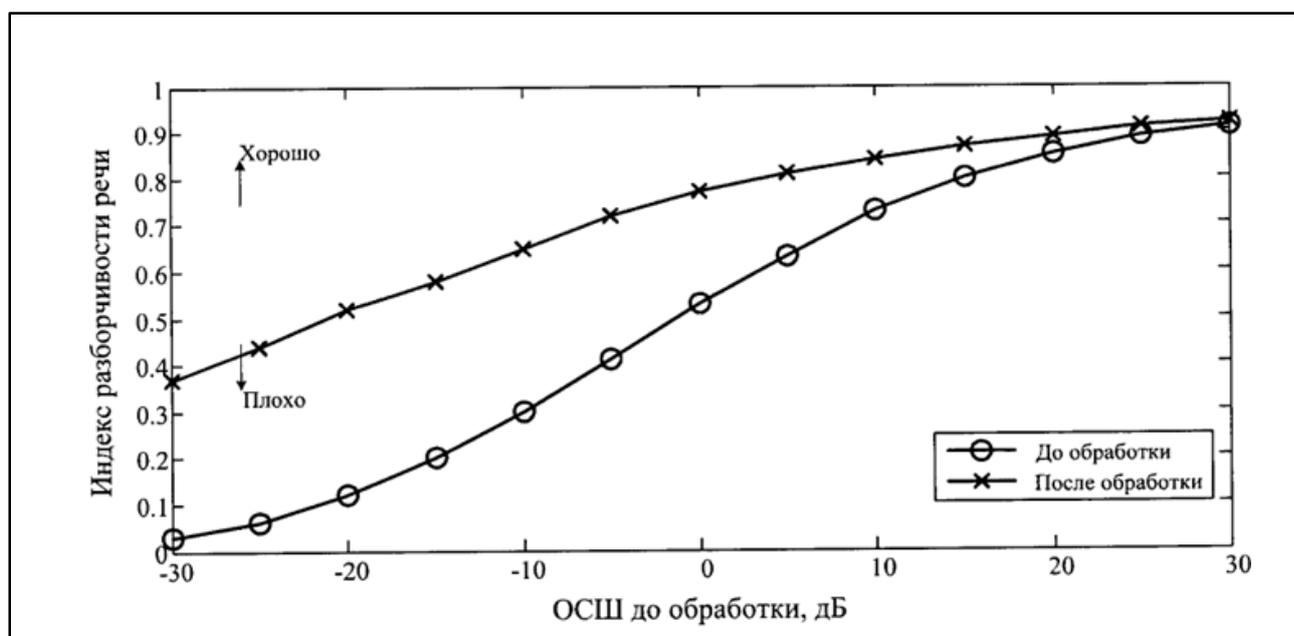


Рисунок 44. Сравнение среднего значения индекса разборчивости СИ для речевых сигналов до и после обработки предложенным способом

В агрессивной шумовой обстановке (ОСШ < -20 дБ) предложенный способ обеспечивает удовлетворительную разборчивость, а при снижении интенсивности шума (ОСШ > 0 дБ) - хорошую. Необработанный сигнал имеет заметно меньшую разборчивость для ОСШ < 0 дБ и становится неразборчивым уже при ОСШ < -10 дБ.

При расчете среднего значения индекса разборчивости SII анализируется эквивалентный уровень спектра речевого сигнала $x[n]$ и эквивалентный уровень спектра шумового сигнала $v[n]$ в каждой из критических полос, вносящих вклад в разборчивость. Эквивалентный уровень рассчитывается, как усредненная по времени мощность сигнала в критической полосе, поделенная на ширину полосы. Поскольку в рассмотренном способе отдельно доступна обработка полезного сигнала $x[n]$ и шумового сигнала $v[n]$, то SII можно вычислить достаточно легко. Величина SII принимает значение от 0 до 1. Если SII больше 0,75, то разборчивость считается хорошей, если же SII имеет значение меньше 0,45, то воспринимаемый слушателем сигнал недостаточно разборчив [74-76].

Кроме того, был разработан и проведен эксперимент, позволяющий оценить повышение разборчивости полезного сигнала $x[n]$ другого типа в агрессивной шумовой обстановке. Для проведения эксперимента были заданы следующие условия:

- полезный сигнал $x[n]$ является музыкальной композицией (песней), прослушиваемой в автомобиле с открытыми окнами;
- шумовой сигнал $v[n]$ акустической обстановки - все шумы от дороги, ям, торможения, работы жесткой подвески.

Эксперимент показал, что в выходном сигнале $v[n]$ при использовании разработанного способа удаляются все шумы, при этом окружающий шум постепенно перестает восприниматься слуховым аппаратом пользователя. Громкость возрастает плавно, звуковое давление комфортное, на слуховой аппарат пользователя изменение громкости значительного влияния не оказывает. При кратковременном резком увеличении окружающего шума, сбоя алгоритма не происходит, ситуация обрабатывается достаточно плавно, скрывая резкое возрастание шума за счет подстройки частот шумового сигнала $v[n]$ и соответственного увеличения громкости полезного сигнала $y[n]$. Создается полное впечатление прослушивания музыкальной композиции в автомобиле с закрытыми окнами.

Таким образом, эффективность разработанного способа и устройства повышения разборчивости полезного сигнала в агрессивной шумовой обстановке признана удовлетворительной.

4.2 ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ РАЗРАБОТАННОГО СПОСОБА И УСТРОЙСТВА ФОРМИРОВАНИЯ ПЕРСОНАЛИЗИРОВАННЫХ СИГНАЛОВ ДЛЯ КОМПЕНСАЦИИ ПОТЕРИ СЛУХА С ПОМОЩЬЮ РЕСУРСОВ МОБИЛЬНЫХ СЕТЕЙ И УСТРОЙСТВ СВЯЗИ

В 2019 г. была запущено пилотное внедрение мобильного приложения Petralex, реализованного совместно с крупнейшим мобильным оператором в Шри-Ланке компанией «Dialog» на основе предложенных в работе способов [77-81, Приложение Б]. В стране насчитывается более 400 000 людей с нарушениями слуха, и каждый третий взрослый возрастом старше 55 лет может страдать от потери слуха [77].

Были проведены исследования работоспособности данного приложения на группах людей с различными патологиями слуха и опытом использования слуховых аппаратов, которые показывают, что использование приложения, в большинстве случаев, улучшило результаты речевой аудиометрии и тональной аудиометрии в свободном поле. Улучшение результатов пациентов с нарушениями слуха при проведении речевой аудиометрии в тишине в среднем составило 30,3%, при шуме +10 дБ – 24,3%, при шуме +5 дБ – 20,8%.

Применение мобильного приложения «Petralex» [77] повышает понимание получаемой слуховой информации на вербальной основе среди респондентов, не имевших опыта использования слуховых аппаратов, до 75–100%. У респондентов, которые имели опыт использования СА, заметно улучшение разборчивости предоставляемого речевого материала только в варианте ответа

«много» (понимание 75% от всего объема предоставленной слуховой информации). В ситуации личной беседы респондентов, которые имеют опыт использования слуховых аппаратов, с одним человеком в шумном окружении использование приложения показывает лучшие результаты, чем использование специализированного СА. У респондентов с бинауральным нарушением слуха (разницей потери слуха ушей) менее 60% улучшение результатов наиболее выражено (рисунок 45).

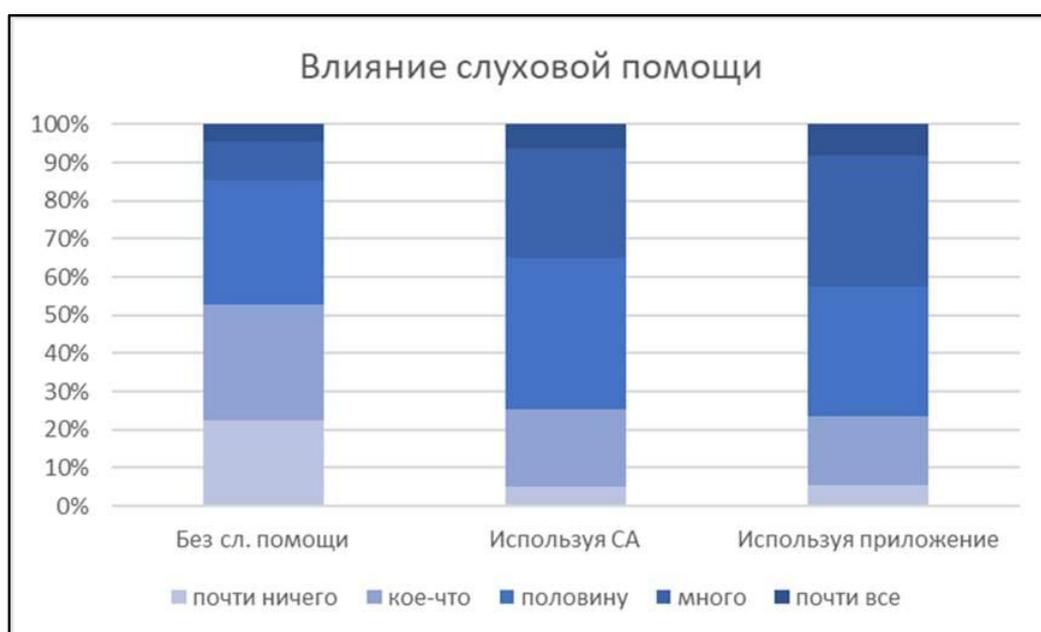


Рисунок 45. Влияние слуховой помощи

Разработанное мобильное приложение предоставляет возможность дистанционной настройки и тестирования слуха специалистом [20-24]. Во время дистанционного сеанса может быть проведено:

- обследование состояния слуха,
- демонстрация функций коррекции,
- настройка средства коррекции по результатам обследования, персонализация настроек,
- инструктаж по использованию средства коррекции,

- введение в курс сурдологической адаптации для ускорения привыкания к звуку и т.д.

Возможность удаленного общения со специалистом облегчает для пациента процесс настройки и эксплуатации средства коррекции, снимает когнитивный диссонанс, относительно правильности/верности настройки, позволяет сохранить и отслеживать в дальнейшем результаты обследования, экономит время, которое ему пришлось бы потратить на физическое посещение аудиолога.

Определенный экономический эффект такое устройство дает и для системы здравоохранения. Такие расчеты не проводились в рамках работы, но параметрами эффективности выступают:

- уменьшение нагрузки на аудиолога за счет автоматизации процесса дистанционного тестирования, расшифровки и настройки;
 - сокращение времени ожидания и проведения приема у специалиста-аудиолога на 90%;
 - снижение расходов на оказание помощи пациентам – на 90%;
 - повышения качества оказания медицинских услуг за счет проведения регулярных осмотров и мониторинга состояния слуха пациентов – на 40%.
- Качество самого процесса тестирования состояния слуха не зависит от формы проведения – очного или дистанционного и при проведении эксперимента показывает одинаковые результаты.

Автоматизация процесса тестирования и расшифровки аудиограммы позволит проводить обучение большого числа консультантов-аудиологов, которые обучены работе с автоматизированной системой и настройке ограниченного списка средств коррекции. Это критично важно для стран с малым количеством специалистов и большим процентом населения нуждающимся в слуховой помощи, а также со слаборазвитой транспортной системой.

Таким образом, эффективность предложенных решений по компенсации потери слуха с помощью мобильных телефонных аппаратов признана хорошей.

4.3. ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ФУНКЦИОНИРОВАНИЯ РЕАЛИЗОВАННОГО УСТРОЙСТВА НА ОСНОВЕ СПОСОБА ФОРМИРОВАНИЯ БАЗЫ ДАННЫХ СБАЛАНСИРОВАННЫХ СИНТАГМ

Разработанное устройство автоматизированного распознавания и преобразования устной речи в текст с учетом модификации и текстонезависимой конверсии голоса с помощью текстового процессора в режиме реального времени, направлено на распознавание активного лексического словарного запаса современного взрослого среднестатистического человека. Лексический запас – количество слов, которые человек понимает и использует в речи. Лексический словарь состоит из двух категорий:

- активный словарный запас – слова, которые человек употребляет регулярно в устной речи и письме;
- пассивный словарный запас – состоит из слов, которые человек знает и воспринимает на слух, но сам не использует. Значительно превышает активный словарь.

По последним данным исследований, активный запас взрослого человека составляет от 5 тыс. до 35 тыс. слов, пассивный – от 20 тыс. до 100 тыс. слов, что в среднем составляет 50% от лексического словарного запаса по академическому словарю русского языка (примерно 130 тыс. слов) [107].

Результатом использования этого устройства на базе разработанного способа формирования сбалансированных синтагм является уменьшение объемов базы данных, необходимой для автоматизированного распознавания устной речи на 25-50%, что приводит к уменьшению вычислительных и информационных ресурсов минимум на 25%. Для оценки эффективности функционирования реализованного устройства необходимо определить, влияет ли уменьшение вычислительных и информационных ресурсов на точность распознавания устной речи.

Для исследования эффективности функционирования разработанного программно-аппаратного комплекса использовался метод экспериментального исследования. На сегодняшний день, нет единых стандартных требований к точности распознавания, особенно в режиме реального времени, обычно ориентируются на оценки точности ведущих в этой области компаний. Хорошей (достаточной) является точность выше 90%.

Для оценки точности распознавания нет стандартных методик, будем использовать методику WER — частота ошибок в словах фразы. Метод использует расстояние Левенштейна - оценка разницы между двумя строками, который учитывает количество вставок, удалений и замен в одной из строк относительно другой на уровне слов. То есть сравнивают по две строки: результат распознавания и исходную строку. При проведении оценки в итоге принято считать две величины:

- Average WER_SR (sum of relationships—сумма отношений) — средняя частота ошибок в строке;
- Average WER_RS (relationship of sums) — средняя частота ошибок в словах.

Формула расчёта:

$$\text{AverageWER_SR} = (S+D+I)/N \quad (4.1)$$

$$\text{AverageWER_RS} = (S+D+I)/C, \quad (4.2)$$

где:

S - количество замен;

D - количество удалений;

I - количество вставок;

C - количество корректных слов;

N — количество слов в исходной строке.

Для точности самой оценки необходимо проанализировать большое количество элементов (аудиозаписей) в датасете, в этом случае формулы 4.1 и 4.2 будут выглядеть следующим образом;

$$\text{Average WER}_{SR} = ((S_1 + D_1 + I_1)/N_1 + \dots + (S_m + D_m + I_m)/N_m)/m \quad (4.3)$$

$$\text{Average WER}_{RS} = (S_1 + \dots + S_m + D_1 + \dots + D_m + I_1 + \dots + I_m)/(C_1 + \dots + C_m), \quad (4.4)$$

где m — количество элементов (аудиозаписей) в датасете.

Для проведения оценки на вход подавался поток устной речи на бытовые темы, без использования профессиональных и специфических тем и не содержащие не словарные слова. В тестовом датасете использовались записи телефонных разговоров. В данных телефонии есть подкатегории мужских и женских голосов, каждая из которых, в свою очередь, делится на короткие (менее 5 слов), средние (от 5 до 10 слов) и длинные записи (более 10 слов). В общей сложности были собраны примерно по 180-200 образцов записей на каждую категорию. Датасет размечен вручную по правилам, описанным выше. Тестовые данные не использовались для обучения нейросетей.

Сама методика распознавания и значения AverageWER , $\text{relationshipofsums}$ и AverageWER , $\text{sumofrelationships}$, взятые в качестве эталонных, была выбрана из открытых источников, чтобы иметь цифры для сравнения, которые были получены при похожих исследованиях [108,109] компаний – лидеров в работе с CAPT, чьи продукты характеризуются точностью от 90%:

- Центр речевых технологий (ЦРТ);
- Microsoft (Азура - azure);
- Яндекс;
- Google;
- Тинькофф (tinkoff);
- «Наносемантика» (ashmanov_8_2).

В таблице 10 приведены используемые эталонные значения оценки точности распознавания по мужской и женской речи при $m=132$ и полученные при проведении эксперимента значения оценки точности разработанного устройства.

Результаты исследования оценки точности, как видно из таблицы 10, уже показывают хорошие результаты точности распознавания, а накопление БД

эталонных синтагм, которая позволит увеличить точность распознавания, продолжается.

Таблица 10. Результаты исследований точности распознавания речи

Номер п/п	Название компании/продукта	Предложения 5-10 слов			
		Мужские голоса		Женские голоса	
		Сумма отношений	Отношение суммы	Сумма отношений	Отношение суммы
1.	Tinkoff	10,39	10,55	12,27	12,29
2.	ashmanov_8_2	12,01	11,89	11,57	11,65
3.	Предлагаемое решение	12,08	12,06	12,8	12,67
4.	ЦРТ	14,32	14,27	14,48	14,67
5.	Яндекс	15,6	16,03	16,69	16,64
6.	Microsoft (Азур)	31,73	31,75	22,12	21,47
7.	Google	45,06	45,09	42,1	40,92

Для оценки используемых ресурсов рассмотрим такой параметр, как оценка времени, необходимого для обучения системы автоматизированного распознавания речи. Эффективным обучением является обучение системы с наилучшим качеством (отсутствие артефактов, естественность речи, хорошая разборчивость) при наименьшей длительности процесса обучения. Как показали испытания, в режиме обучения этого устройства вместо зачитывания текста из 100 фраз диктору нужно прочесть всего 60÷75 фраз, соответствующих сбалансированным синтагмам, что с хорошей точностью распознавания сокращает время обучения системы не менее чем на 25%.

Также при проведении эксперимента по распознаванию устной речи были выявлены устойчивые признаки уменьшения времени обработки звуковых сигналов в среднем на 15%, что в первую очередь, связано с уменьшением объемов БД и, соответственно, количеством итераций обучения. На данный момент количество запусков распознавания является недостаточным для объективной оценки, так как процесс накопления БД фонем, и синтагм занимает достаточно большое время. Тем не менее, результаты тестовых прогонов

показывают достаточную точность при значительном уменьшении требуемых ресурсов и увеличении скорости обработки на 15%.

Выводы по главе 4.

1. Исследования эффективности всех предложенных технических решений проводились с помощью проведения экспериментов. Для каждого конкретного эксперимента выбиралась стратегия, критерии эффективности и, при необходимости, создавалось специальное программное обеспечение.

2. Выбор экспериментального метода оценки эффективности обусловлен практической направленностью темы и цели диссертационной работы – разработать комплексную ассистивную технологию для увеличения доступа слабослышащих абонентам к ресурсам систем, сетей и устройств телекоммуникации и требовал практических испытаний.

3. Эксперименты показали, что при внедрении разработанного устройства повышения разборчивости полезного сигнала в агрессивной шумовой обстановке повышается индекс разборчивости SII от удовлетворительного до хорошего в зависимости от уровня окружающего шума и степени потери слуха. Для полезного сигнала – музыкальной композиции удаляются все шумы, при этом окружающий шум постепенно перестает восприниматься телефоном пользователя. Громкость возрастает плавно, звуковое давление комфортное, на телефон пользователя изменение громкости значительного влияния не оказывает.

4. Использование устройства формирования персонализированных сигналов увеличивает функционал мобильного телефона не только возможностью совмещения его со слуховым аппаратом, но и функциями дистанционной настройки возможностей СА в мобильном телефоне и дистанционного тестирования слуха, что улучшило результаты речевой аудиометрии и тональной аудиометрии в среднем составило 30,3%, при шуме +10 дБ – 24,3%, при шуме +5 дБ – 20,8%.

5. Проведение эксперимента по эффективности функционирования разработанного устройства АРР несколько затруднено на данном этапе, так как в настоящее время идет накопление БД сбалансированных синтагм при работе с разными дикторами и текстами устной речи в базе данных для дальнейшего обучения нейронной сети, но выполненный эксперимент показывает, что время обучения уменьшается значительно – до 25% при уменьшении времени обработки сигнала и показателях, обеспечивающих удовлетворительную точность. Информационный ресурс, необходимый для распознавания устной речи – БД сбалансированных синтагм при использовании предложенного метода по расчетам может уменьшиться до 30%.

Таким образом, можно считать, что все предложенные научно-практические и технические решения, направленные на достижение поставленной в диссертационной работе цели, были решены с достаточной степенью эффективности.

ЗАКЛЮЧЕНИЕ. В рамках выполнения диссертационной работы была разработана комплексная ассистивная технология, обеспечивающая клиенто-ориентированный доступ слабослышащих абонентов к ресурсам сетей и систем телекоммуникаций за счет разработанных и реализованных программно-аппаратных комплексов, повышающих разборчивость и распознавание звуковых сигналов с помощью мобильных устройств в условиях агрессивной шумовой обстановки окружающей среды, таких как:

1. Блок обработки звуковых сигналов с адаптивным компрессором динамического диапазона, которым изменяют динамический диапазон полезного сигнала. При обработке усиливаются лишь те фрагменты полезного сигнала, которые маскируются шумом. В полной тишине в полезный сигнал не вносятся никаких изменений (Патент РФ RU 2589298 C1). Реализованное устройство позволяет повысить естественность, разборчивость и информативность звучания звуковых сигналов, в том числе и музыкальных композиций при прослушивании с помощью мобильного устройства связи или слухового аппарата в акустической шумовой обстановке за счет использования в алгоритме функционирования устройства разработанного способа снижения эффекта маскирования полезного звукового сигнала нестационарными акустическими шумами при помощи использования частотно-зависимого адаптивного усиления. Проведенное исследование эффективности показало повышение индекс разборчивости SII на 30-40% в зависимости от уровня шума (чем выше шум, тем больше повышение индекса разборчивости).

2. Программно-аппаратный комплекс, позволяющий слабослышащему абоненту использовать мобильный телефон в трех режимах:

- режим слухового аппарата для прослушивания мультимедийных устройств,
- режим слухового аппарата для разговора с человеком-собеседником, находящимся рядом со слабослышащим пользователем,
- режим телефонного аппарата для слабослышащих,

- режим системы автоматизированного распознавания и преобразования речи в текст в реальном времени с выводом транскрибированной речи на экран мобильного устройства, а также обеспечивающее дистанционные настройку телефона и диагностирование слуха слабослышащих абонентов [патент Китая 10155840716, Патент РФ RU 2568281 C2, патент WO WO/2014/193,264). Для реализации такого комплекса был разработан способ компенсации потери слуха за счет формирования персонализированных аудиосигналов для слабослышащих пользователей на основе их атрибутов, хранящихся в базе данных на сервере сети связи и имеющих привязку к телефонным номерам. Все разработанные приложения устройства компенсации потери слуха позволяют слабослышащему абоненту удобно и эффективно, с учетом индивидуальных особенностей, использовать мобильный телефон для доступа к ресурсам цифровых сетей. Проведенные во время исследования эффективности эксперименты показали, что улучшение результатов пациентов с нарушениями слуха при проведении речевой аудиометрии в тишине в среднем составило 30,3%, при шуме +10 дБ – 24,3%, при шуме +5 дБ – 20,8%. У респондентов с бинауральным нарушением слуха (разницей потери слуха ушей) улучшение результатов наиболее выражено - более 60%.

3. Программно-аппаратный комплекс транскрибирования устной речи на основе разработанного способа формирования базы данных сбалансированных синтагм посредством текстового процессора с применением метода транскрипционного моделирования, обеспечивающего хорошую степень точности распознавания, увеличение скорости обработки при уменьшении количества информационных ресурсов за счет целевого обучения нейронной сети наиболее часто используемыми фонемами, аллофонами и синтагмами наиболее часто встречающимися в устной речи (Патенты РФ RU 2460154 C1, РФ RU 2591640 C1, РФ RU 2510954 C2, РФ RU 2427044 C1, WO/2013/180,600, WO/2012/173,516).

4. Разработанные в диссертационной работе имеют научную и практическую значимость для обеспечения доступа к ресурсам систем, сетей и устройств телекоммуникации не только слабослышащих абонентов, но и представляют научное обоснование и практическую возможность расширения функционала мобильных устройств, обеспечения индивидуального клиенто-ориентированного подхода с использованием современных информационных технологий для всех абонентов систем, сетей телекоммуникации.

Таким образом, все поставленные в диссертационной работе задачи были решены, эффективность предложенных решений доказана, цель достигнута.

СПИСОК ЛИТЕРАТУРЫ

1. Digital 2024: Основные выводы из ежегодного отчета DataReportal <https://www.byyd.me/ru/blog/2024/02/digital-2024-datareportal/>
2. Российский рынок AssistiveTech: особенности и перспективные управления: Агентство инноваций Москвы: https://portal.inno.msk.ru/uploads/agency-sites/analytics/research/AssistiveTech+in+Russia_AIM_2023.pdf/
3. Патент № 2589298 Российская Федерация, МПК G10L 19/02 G10L 21/232 G10L21/28 C1. Способ повышения разборчивости и информативности звуковых сигналов в шумовой обстановке: № 2014153295/08: заявл. 29.12.2014: опубл. 10.07.2016 / Вашкевич М.И, Азаров И.С., Бредихин А.Ю., Петровский А.А.: заявитель Бредихин А.Ю. – 20 с.: ил. – Текст непосредственный.
4. Патент № 2568281 Российская Федерация, МПК H04W 4/18 (2009.01) H04M 1/247 (2006.01). Способ компенсации потери слуха в телефонной системе и в мобильном телефонном аппарате: № 2013125243/08: заявл. 31.05.2013: опубл. 20.11.2015/ Вашкевич М.И, Азаров И.С., Бредихин А.Ю., Петровский А.А.: заявитель Бредихин А.Ю. – 27 с.: ил. – Текст непосредственный.
5. Патент № WO/2014/193264, МПК H04W 4/18 2009.1 H04M 1/247 2006.1 H04R 25/00 2006.1 G10L 21/02 2013.1. Method for compensating for hearing loss in a telephone system and in a mobile telephone apparatus: № PCT/RU2014/000297: заявл. 23.04.2014: опубл. 04.12.2014 / A. Bredikhin, M. Vashkevich, I. Azarov, A. Petrovsky: заявитель / A. Bredikhin, – 27 с.: ил. – Текст непосредственный.
6. Патент № WO/2008/130265 МПК G06F 17/30 2006.1. Метод ввода и поиска информации об объекте в удаленной базе данных: № PCT/RU2007/000295 заявл. 04.06.2007: опубл. 30.10.2008/ Бредихин А.Ю.– 22 с.: ил. – Текст непосредственный.
7. Патент № CN101558407 Китай, МПК G06F 17/30 Method for inputting and searching information about an object in a remote database: № 200780045919: заявл. 04.06.2007: опубл. 14.10.2009 / A. Bredikhin, N. Sergeichev: заявитель / A. Bredikhin, – 22 с.: ил. – Текст непосредственный.
8. Патент № US7813723B2, МПК H04L 12/66 (200601), H04L 29/06 (2006.01) US. C1. 370/352; 455/4143. Method and system for connecting a voice call using a domain name database: № US12/692,616 заявл. 12.10.2010: опубл. 12.10.2010/ A. Bredikhin, N. Sergeichev: заявитель / A. Bredikhin, – 23 с.: ил. – Текст непосредственный.

9. Патент № 2460154 Российская Федерация, МПК G10L 13/08 (2006.01) G06F 17/21 (2006.01). Способ автоматизированной обработки текста и компьютерное устройство для реализации этого способа: № 2011124155/08: заявл. 15.06.2011: опубл. 27.08.2012 / Бредихин А.Ю., Сергейчев Н.Е.: заявитель Бредихин А.Ю. – 21 с.: ил. – Текст непосредственный.
10. Патент № 2591640 Российская Федерация, МПК G10L 13/00(2006.01) G10L 13/033(2013.01). Способ модификации голоса и устройство для его осуществления (варианты): № 2015119825/08: заявл. 27.05.2015: опубл. 20.07.2016 / Вашкевич М.И, Азаров И.С., Лихачев А.Ю., Бредихин А.Ю., Петровский А.А.: заявитель Бредихин А.Ю. – 25 с.: ил. – Текст непосредственный.
11. Патент № 2510954 Российская Федерация, МПК G10L 13/00(2006.01) Способ переозвучивания аудиоматериалов и устройство для его осуществления: № 2012120562/08: заявл. 18.05.2012: опубл. 10.04.2014 / Бредихин А.Ю.: заявитель Бредихин А.Ю. – 25 с.: ил. – Текст непосредственный.
12. Патент № 2427044 Российская Федерация, МПК G10L21/00. Текстозависимый способ конверсии голоса: № 2010119144/09А: заявл. 14.05.2010: опубл. 20.08.2011/ Бредихин А.Ю., Петровский А.А., Сергейчев Н.Е.: заявитель Закрытое акционерное общество "Ай-Ти Мобайл" (РФ)– 21 с.: ил. – Текст непосредственный.
13. Патент № WO/2013/180600, МПК G10L 13/02 2013.1 G10L 15/07 2013.1. Method for rerecording audio materials and device for the implementation thereof: № PCT/RU2013/000404: заявл. 16.05.2013: опубл. 05.12.2013/ А. Bredikhin: заявитель / А. Bredikhin, – 27 с.: ил. – Текст непосредственный.
14. Патент № 2393548 Российская Федерация, МПК G10L 13/00 (2006.01). Устройство для изменения входящего голосового сигнала в выходящий голосовой сигнал в соответствии с целевым голосовым сигналом: № 2008147022/09: заявл. 28.11.2008: опубл. 27.06.2010/ Бредихин А.Ю., Сергейчев Н.Е.: заявитель Общество с ограниченной ответственностью "Конвент Люкс" (РФ) – 21 с.: ил. – Текст непосредственный.
15. Патент № 66103 Российская Федерация, МПК G10H 1/36(2006.01) G06F 17/30(2006.01) G10L 13/00(2006.01). Устройство обработки речевой информации для модуляции входного голосового сигнала путем его преобразования в выходной голосовой сигнал: № 2007118609/22: заявл. 21.05.2007: опубл. 27.08.2007/ Бредихин А.Ю., Сергейчев Н.Е.: заявитель

Общество с ограниченной ответственностью "ТЕЛЕКОНТЕНТ" (РФ) – 25 с.: ил.
– Текст непосредственный.

16. Патент № WO/2012/173516, МПК G10L 13/08 2013.1 G06F 17/21 2006.1. Method and computer device for the automated processing of text № PCT/RU2012/000328 заявл. 26.04.2012: опубл. 20.12.2012 / А. Bredikhin, N. Sergeichev: заявитель / А. Bredikhin, – 23 с.: ил. – Текст непосредственный.
17. Патент № MYPI 20092785 Малазия, МПК G06F 17/30. Method for inputting and searching information about an object in a remote database: № PI 20092785: заявл. 04.06.2007: опубл. 23.10.2008/ А. Bredikhin, N. Sergeichev: заявитель / А. Bredikhin, – 21 с.: ил. – Текст непосредственный.
18. Патент № IN2995/DELNP/2009 Индия, МПК G06F 17/30. Method for inputting and searching information about an object in a remote database: № 2995/DELNP/2009: заявл. 05.05.2009: опубл. 01.01.2010 / А. Bredikhin, N. Sergeichev: заявитель / А. Bredikhin, – 21 с.: ил. – Текст непосредственный.
19. Патент № EP2141611 Европейское патентное ведомство (ЕПВ), МПК G06F 17/30 G06Q 30/02. Method for inputting and searching information about an object in a remote database: № 07834952: заявл. 04.06.2007: опубл. 06.01.2010 / А. Bredikhin, N. Sergeichev: заявитель / А. Bredikhin, – 21 с.: ил. – Текст непосредственный.
20. Патент № US20090259640 США, МПК H04L 29/06 G06F 15/16 G06F 17/30 G06Q 30/00 H04M 1/72 H04M 1/725. Method for inputting and searching information about an object in a remote database: № 12493956: заявл. 29.06.2009: опубл. 15.10.2009 / А. Bredikhin, N. Sergeichev: заявитель / А. Bredikhin, – 21 с.: ил. – Текст непосредственный.
21. Патент № MX/a/2009/011194 Мексика, МПК C04B 111/20 C04B 111/40 C04B 28/26 C04B 38/00 C04B 40/02 G06F 17/30. Method for inputting and searching information about an object in a remote database: № MX/a/2009/011194: заявл. 16.10.2009: опубл. 17.05.2010 / А. Bredikhin, N. Sergeichev: заявитель / А. Bredikhin, – 21 с.: ил. – Текст непосредственный.
22. Патент № KR1020090113303 Республика Корея, МПК G06F 17/30 H04W 4/02. Method for inputting and searching information about an object in a remote database: № 1020097017404: заявл. 04.06.2007: опубл. 29.10.2009/ А. Bredikhin, N. Sergeichev: заявитель / А. Bredikhin, – 21 с.: ил. – Текст непосредственный.
23. Патент № CA2674110 Канада, МПК G06F 17/30. Method for inputting and searching information about an object in a remote database: № 2674110: опубл.

- 30.10.2008/ A. Bredikhin, N. Sergeichev: заявитель / A. Bredikhin, – 21 с.: ил. – Текст непосредственный.
24. Патент № AU2007351918B2, МПК G06F16/951. Method for inputting and searching information about an object in a remote database: № AU2007351918A. заявл. 24.03.2011: опубл. 24.03.2011 / A. Bredikhin, N. Sergeichev: заявитель / A. Bredikhin, – 20 с.: ил. – Текст непосредственный.
25. Абдюханов, Р.О. Применение технологии широкополосного изменения огибающей спектра звукового сигнала Petralex® в онлайн-приложениях коррекции слуха /Р.О. Абдюханов, А.Ю. Бредихин, Д.Ю. Динцис, А.А. Петровский и др. // Системный администратор. - №7, 2014. – С.136-138
26. Бредихин, А.Ю Система дистанционной настройки средств коррекции слуха/ Паньшин И.А., Порхун М.И., Богати С.Р.: Проектирование и технологии электронных средств № 12 – стр.16-20
27. Dintsis, D. Virtual Learning for People with Hearing Impairs/D. Dintsis, A. Bredikhin // IEEE AET Annual 2015 Conference. Naples, Italy. Publisher: IEEE
28. Бредихин, А.Ю. Способ дистанционной настройки устройства коррекции слуха / А.Ю. Бредихин, И.А. Паньшин, С.Р. Богати // LXIII International Scientific-Practical Conference «EurasiaScience». – 2024. - г. Москва. – С.31-38.
29. Бредихин, А.Ю. Метод дистанционной настройки средств коррекции слуха / А.Ю. Бредихин, И.А. Паньшин, С.Р. Богати // XXVII Международная научно-практическая конференция «Наука и технологии: модернизация, инновации, прогресс. - 29 июля 2024 года, г. Анапа. – С. 9-17.
30. Бредихин, А.Ю. Разработка аппаратно-программного комплекса цифрового устройства звукоусилителя/ А.Ю. Бредихин, А.Г. Самойлов// Сборник научных статей VIII Международной научно-практической конференции «Актуальные проблемы и перспективы развития радиотехнических и инфокоммуникационных систем» «Радиоинфоком-2024» стр. 407-411
31. Бредихин, А.Ю. «Система дистанционной настройки средств коррекции слуха»/А.Ю. Бредихин, И.А. Паньшин, М.И. Порхун, С.Р. Богати// XVI Международная Научная Конференция «Физика и Радиоэлектроника в Медицине и Экологии ФРЭМЭ 2024» 02-04 июля 2024, стр.418-422
32. Патент № 2324296 Российская Федерация, МПК H04M 1/64(2006.01) H04W 4/14(2009.01) H04W 4/18(2009.01). Способ обмена сообщениями и устройства для его реализации: № 2007110749/09: заявл. 26.03.2007: опубл. 10.05.2008/

- Бредихин А.Ю., Сергейчев Н.Е.: заявитель Закрытое акционерное общество "Ай-Ти Мобайл" (РФ) – 25 с.: ил. – Текст непосредственный.
33. Патент № 2393548 Российская Федерация, МПК G10L 13/00(2006.01). Устройство для изменения входящего голосового сигнала в выходящий голосовой сигнал в соответствии с целевым голосовым сигналом: № 2008147022/09: заявл. 28.11.2008: опубл. 27.06.2010/ Бредихин А.Ю., Сергейчев Н.Е.: Общество с ограниченной ответственностью "Конвент Люкс" (РФ) – 25 с.: ил. – Текст непосредственный.
34. Патент № WO/2008/118038, МПК H04M 1/64 2006.1 G10L 13/08 2006.1 H04W 4/18 2009.1 H04W 4/14 2009.1. Message exchange method and devices for carrying out said method: № PCT/RU2007/000638 заявл. 19.11.2007: опубл. 2.10.2008 / А. Bredikhin, N. Sergeichev: заявитель / А. Bredikhin, – 23 с.: ил. – Текст непосредственный.
35. Deafness and hearing loss [Электронный ресурс] / ВОЗ. – URL: <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>.
36. Hearing Loss Facts and Statistics [Электронный ресурс] / Hearing Loss Association of America (HLAA) /URL: https://www.hearingloss.org/wp-content/uploads/HLAA_HearingLoss_Facts_Statistics.pdf
37. Бредихин, А.Ю. Интервью с вице-президентом ААР Китая Bian Weiguo <https://www.ncoa.org/adviser/hearing-aids/low-hearing-loss-treatment-reasons>
38. Petralex: [Электронный ресурс] / Dialog Foundation. – URL: <https://dialogfoundation.org/sustainability-initiatives/digital-inclusion/petralex/>.
39. Андреев, Г. А., Самойлов А. Г., Самойлов С. А. Схемотехника устройств формирования сигналов: учебное пособие <http://e.lib.vlsu.ru:80/handle/123456789/1312>
40. ГОСТ Р ИСО 8253-1-2012. Акустика. Методы аудиометрических испытаний. Тональная пороговая аудиометрия по воздушной и костной проводимости. Часть 1: Национальный стандарт Российской Федерации: дата введения 2013-12-01 / Федеральное агентство по техническому регулированию и метрологии от 29 ноября 2012 г. N 1386-ст.
41. Yanz J.L. Phone and hearing aids: issues, resolutions, and a new approach. *Hearing Journal*, 2005, 58 (10), pp.41-48

42. Dong R., Hermann D., Brennan R., Chau E., Joint filter bank structures for integrating audio coding into hearing aid applications. - ICASSP- 2008, - pp.1533-1536).
43. American National Standard. Methods for the Calculation of the Speech Intelligibility Index. ANSI S3.5-1997, 1997
44. Галкин, В.А. Цифровая мобильная радиосвязь: учебн. пособие. – М. : Горячая линия – Телеком, 2021. – 518 с
45. Волков, А.А., Карпова Г.В., Журавлев О.Е. Повышение помехоустойчивости радиосвязи / А.А. Волков, Г.В. Карпова, О.Е. Журавлев // Мир транспорта, 2012. – №3. – С. 31–33
46. Горошко, С.М., Петров С.Н. Метод шумоочистки речевых сигналов на основе мел-частотных кепстральных коэффициентов с использованием фильтрации Калмана / Известия Гомельского государственного университета имени Ф. Скорины, № 6 (117), 2019
47. Завьялов, С.В. О возможности увеличения длительности оптимальных сигналов в условиях сохранения символьной скорости передачи / С.В. Завьялов, А.С. Овсянникова // Доклады 19 Международной научно-технической конференции «Цифровая обработка сигналов и её применение» - DSPA-2017. – М.: РНТО РЭС им. А.С. Попова. -Том 1. – С. 56-61.
48. Ключев, Л.Л. Теория электрической связи: учебник. – М.: НИЦ ИНФРА-М, 2019. – 448 с
49. Санников, В.Г. Теория информации и кодирования: учебн. пособие / Санников В.Г. М.: Московский технический университет связи и информатики, 2015-95с
50. Санников, В.Г. Адаптивный оптимальный интеллектуальный модем для системы цифровой телефонии / В.Г. Санников // ЭЛЕКТРОСВЯЗЬ. – 2014. – № 5. – С. 26-29
51. Kates J.M., Arehart K.H., Multichannel dynamicrange compression using digital frequency warping: EURASIP J. Adv. Sig. Proc, vol. 2005, no. 18, pp.3003-3014, 2005
52. Кузин, Л. Т. Основы кибернетики: В 2-х томах.– М: Энергия, 1979. — 576с.

53. Кислицын, А.Б. Формирование и прием спектрально-эффективных многочастотных сигналов с неортогональным частотным уплотнением на основе БПФ/ОБПФ уменьшенной размерности / А.Б. Кислицын, А.В. Рашич // Электромагнитные волны и электронные системы № 7, т. 19, 2014 - М.: ЗАО «Издательство «Радиотехника», 2014. – 46-53 с.
54. Крейнделин, В.Б. Новые методы обработки сигналов в системах беспроводной связи / В.Б. Крейнделин // – СПб: Линк, 2009. – 276 с
55. Леньшин, А.В. Помехоустойчивость приема спектрально эффективных сигналов с неортогональным частотным уплотнением / А.В. Леньшин, Е.
56. Макаров, С.Б. Применение алгоритма Витерби для повышения помехоустойчивости приема клиппированных сигналов с модуляцией на одной несущей / С.Б. Макаров, А.М. Марков // Доклады 16 Международной научно-технической конференции «Цифровая обработка сигналов и её применение» - DSPA-2014. – М.: РНТО РЭС им. А.С. Попова. - Том 1. – С. 105-109
57. Макаров, С.Б. Оптимизация формы огибающей спектрально-эффективных многочастотных сигналов / С.Б. Макаров, С.В. Завьялов // Радиотехника. «Электромагнитные волны и электронные системы». – 2014. – Т. 19. - № 7. – С. 38 – 45
58. Paz Martinez-Beneyto, Sebastiano Franchella, Fabio Alonso Rodriguez¹, Rafael Navarro-Velasquez, Miguel A. Martinez-Beneito, Alessandro Martini, Jaime Marco Algarra Are smartphone applications (App) useful to improve hearing? // Acta otorhinolaryngologica italica 2020 (№40). – стр. 304-310
59. Zorila T.-C. Speech-in- noise intelligibility improvement based on spectral shaping and dynamic range compression / T.-C. Zorila, V. Kandida, Y. Stylianou // In Proc. Interspeech, 2012. - Portland, Oregon, 2012. - P. 635-638.
60. Морозов, В. В., Ланцов В. Н., Самойлов А. Г. Труды Владимирского государственного университета: вып. 1: Информационно-телекоммуникационные технологии и электроника.

61. Родионов, А.Ю. Многочастотные цифровые системы связи в условиях многолучевого распространения и их энергетическая эффективность / А.Ю. Родионов // Вестник ДВО РАН. - 2007. - №1. - С. 69 - 72.
62. Самойлов, А. Г., Самойлов С. А. Устройства генерирования и формирования сигналов: учебное пособие- <http://e.lib.vlsu.ru:80/handle/123456789/7372>
63. Самойлов, А. Г., Самойлов С. А. Устройства электропитания: учебное пособие <http://dspace.www1.vlsu.ru/handle/123456789/8901>
64. Санников, В.Г. Цифровая передача непрерывных сообщений на основе дифференциальной импульсно-кодовой модуляции: учебное пособие / В.Г. Санников // – М.: Горячая линия – Телеком, 2016. – 98 с
65. Boothroyd A., Fitz K., Kindred J., etc. Hearing aids and wireless technology, The Hearing Review, March 2008
66. Феер, К. Беспроводная цифровая связь. Методы модуляции и расширения спектра / К. Феер // Пер. с англ. под ред. В.И. Журавлева. - М.: Радио и связь, 2000. –520 с
67. Тихвинский, В.О. Сети мобильной связи LTE. Технологии и архитектура / В.О. Тихвинский, С.В. Терентьев, А.Б. Юрчу // – М.: Эко - Трендз, 2010. – 284 с.
68. Сидоренко, А.А. Мягкое декодирование блочных кодов с использованием синдромной решетки // Перспективные технологии в средствах передачи информации: Материалы 14-ой международной научно – технической конференции. – Владимир: ВлГУ. – 2021. – С. 184–187.
69. Скляр, Б. Цифровая связь. Теоретические основы и практическое применение / Б. Скляр // пер. с англ. – М.: ИД «Вильямс», 2003. – 1104 с
70. Козлова, С.В., Вашкевич М.И. Применение мобильного приложения «Petralex» для оценки слуховых нарушений /Оториноларингология. Восточная Европа 2018 (Том 8, №1). – стр. 81.
71. Phillips C., Loizou Speech enhancement theory and practice: 1st ed. Boca Raton, FL.: CRC, 2007. Releases Taylor & Francis
72. Proakis J.G., Salehi M. Digital communications. 5th ed., McGraw-Hill, New York, 2008. – 1150 p.

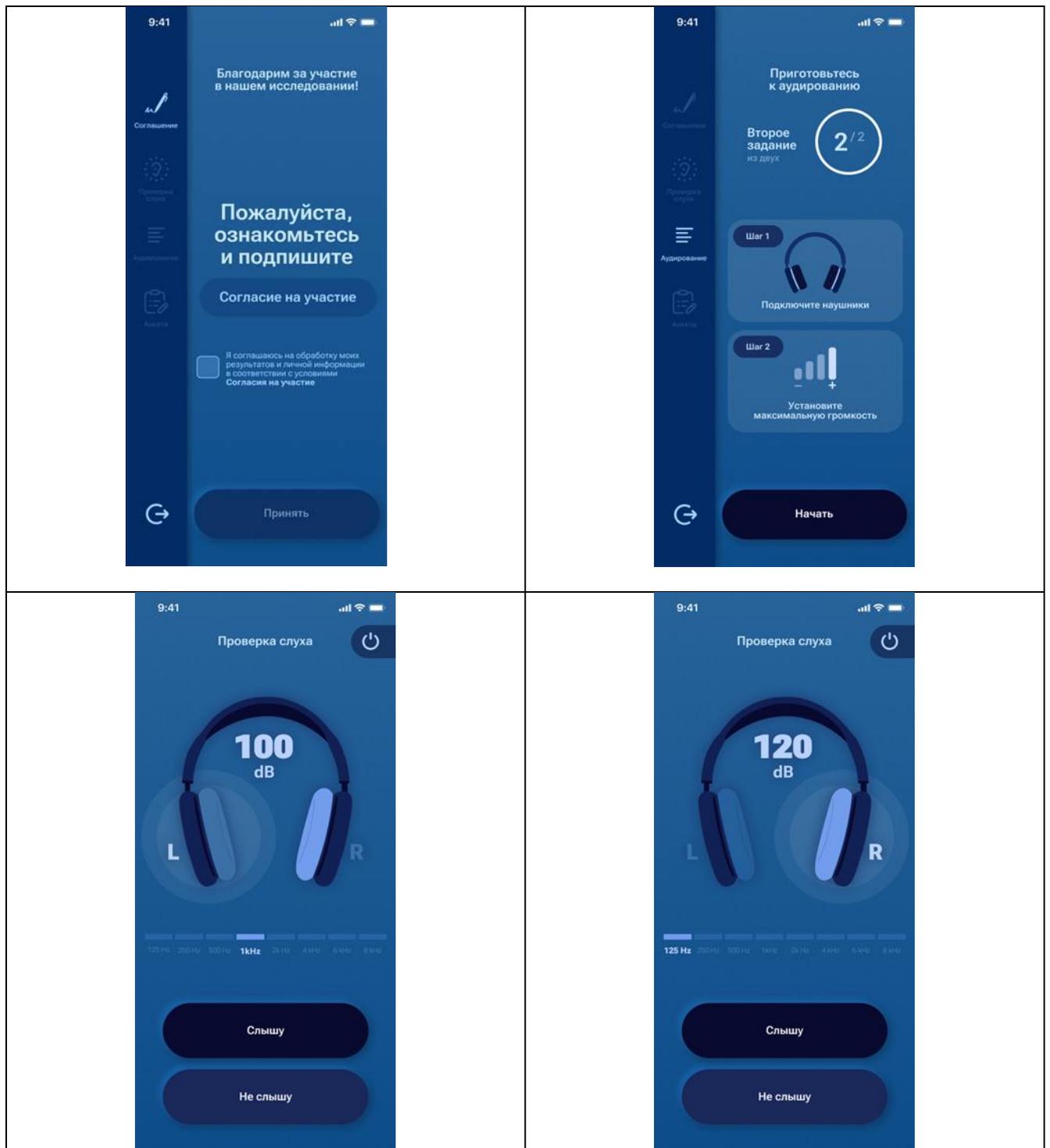
73. Russell J., Niederjohn, James H., Grotelueschen, The enhancement of speech intelligibility in high noise levels by highpass filtering followed by rapid amplitude compression: in Proc. of ICASSP, Aug. 1976, vol. 24, pp. 277-28
74. Sauert B., Vary P., Near end listening enhancement optimized with respect to speech intelligibility: Proc. 17th European Signal Processing Conference (EUSIPCO), pp. 1844-1849, 2009
75. Sauert, B. Near end listening enhancement: speech intelligibility improvement in noisy environments / B. Sauert, P. Vary // ICASSP 2006: proc. of the International Conference on Acoustic, Speech, and Signal Processing. - Toulouse, France, 2006. - P. 493-496
76. Сайт Petralex: <https://petralex.pro/ru/blog/4>
77. Бредихин, А.Ю. Способ дистанционной настройки устройства коррекции слуха / А.Ю. Бредихин, И.А. Паньшин, С.Р. Богати // LXIII International Scientific-Practical conference «EurasiaScience». – 2024. - г. Москва. – С.31-38
78. Бредихин, А.Ю. Метод дистанционной настройки средств коррекции слуха / А.Ю. Бредихин, И.А. Паньшин, С.Р. Богати // XXVII Международная научно-практическая конференция «Наука и технологии: модернизация, инновации, прогресс. - 29 июля 2024 года, г. Анапа. – С. 9-17.
79. Dintsis, D. Virtual Learning for People with Hearing Impairs / D. Dintsis, A. Bredikhin // IEEE AEIT Annual 2015 Conference. Naples, Italy. Publisher: IEEE
80. Бредихин, А.Ю. Система дистанционной настройки средств коррекции слуха / Бредихин А.Ю., Паньшин И.А., Порхун М.И., Богати С.Р. // Проектирование и технология электронных средств - № 12, 2024, стр. 16-20
81. Бредихин, А.Ю. Актуальность применения дистанционных средств диагностики коррекции слуха / Бредихин А.Ю., Самойлов А.Г. // Проектирование и технология электронных средств - № 1, 2025, стр. 2-6
82. Васильев, В. И. Распознающие системы. Справочник. — К.: Наукова думка, 1983.— 422 с.
83. Сеницын, А.С. Изучение возможности применения методов машинного обучения в радиофизических задачах / А.А. Адеркина, А.С. Сеницын, М.В.

- Махлышев // Труды XXV научной конференции по радиофизике: сборник тезисов конференции. – Нижний Новгород, 2021. – С. 228-231.
84. Тюльпа, Д. Применение искусственного интеллекта в промышленности / Д. Тюльпа // САПР и графика. – 2020. – № 1. – С. 26-30
85. Elizarov D.A., Kolpakova P.E. Application of the transcribing system. In: *Matematicheskoe i eksperimental'noe modelirovanie fizicheskikh processov: Vserossijskaja nauchno-prakticheskaja konferencija smezhdunarodnym uchastiem (Birobidzhan, 15 dekabrja 2022 g.) [Mathematical and Experimental Modeling of Physical Processes]* : Collection of materials of the All-Russian scientific and practical conference with international participation, Birobidzhan, December 15, 2022. Ed. by V.M. Kozin. Birobidzhan Sholom-Aleichem Priamur State University, 2023. Pp. 13–16.
86. Остроух, А.В. Системы искусственного интеллекта: монография / Остроух А.В., Суркова Н.Е. – 2-е изд. – Санкт-Петербург: Лань, 2023 -228 с.
87. Остроух, А.В., Суркова Н.Е. Проектирование информационных систем; монография. – СПб: Издательство «Лань», 2021 – 164 с.
88. Остроух, А.В., Николаев А.Б. Интеллектуальные информационные системы и технологии монография. – СПб: Издательство «Лань», 2019 – 308 с.
89. Тампель, И.Б., Карпов А.А. Автоматическое распознавание речи: Учебное пособие. - СПб: Университет ИТМО, 2017. – 152 с.
90. Алперин, Е.Д., Кнох В.Я. Способ компиляционного фонемного синтеза русской речи и устройство для его реализации /Патент RU 2298234
91. Барабаш, Ю. Л., Зиновьев Б. В Вопросы статической теории распознавания. — М.: Сов. радио, 1967. — 400 с.
92. Горелик, О. Л., Скрипкин В. А. Методы распознавания. — 2 изд. — М.: Высшая школа, 1986. — 208 с.
93. Дуда, Р., Харт П. Распознавание образов и анализ сцен. — М.: Мир, 1976. — 521 с.
94. Кипяткова, И. С., Карпов А. А, Кулешов С. В., Зайцева А. А. Методы и модели автоматического распознавания речи: учебное пособие- Санкт-Петербург: ФИЦ РАН, 2021- 115 с.
95. Devyatkina E. Methods of Translating Video into Text, Automatic Transcription. Yagla. URL: <https://yagla.ru/blog/marketing/6-sposobov-perevesti-audio-i-video-v-tekst--2110m94955/>

96. Ibusheva, M. Translation of audio and video into text: methods of transcription // SEOnews. August 1, 2021 URL: <https://www.seonews.ru/analytics/7-sposobov-perevoda-video-v-tekst/>
97. Масловская, А.Г. Численная обработка экспериментальных данных. Фурье- и вейвлет-анализ временных рядов. Практикум / А.Г. Масловская – Благовещенск: Амурский гос. ун-т, 2020. – 55 с.
98. Лобанов, Б.М., Цирульник Л.И. Компьютерный синтез и клонирование речи / Национальная Академия Наук Белоруссии, Объединенный институт проблем информатики, Минск, Белорусская наука, 2008, стр.198-243
99. <http://e.lib.vlsu.ru:80/handle/123456789/1087>
100. Van Nguyen, T. Machine learning for wideband localization / T. Van Nguyen, Y. Jeong, H. Shin [et al.] // IEEE Journal on Selected Areas in Communications. – 2015. – Vol. 33 (7). – P. 1357-1380. – DOI: 10.1109/JSAC.2015.2430191
101. Zappone, A. Modelaided wireless artificial intelligence: embedding expert knowledge in deep neural networks for wireless system optimization / A. Zappone, M. Di Renzo, M. Debbah [et al.] // IEEE Veh. Technol. Mag. – 2019. – Vol. 14 (3). – P. 60-69.
102. Zhang, Y. Path loss prediction based on machine learning: principal method and data expansion / Y. Zhang, J. Wen, G. Yang [et al.] // Appl. Sci. – 2019. – Vol. 9 (9). – P. 1908.
103. 5 ways to transcribe audio to text. MyNewsdesk, 2019. April 18. URL:<https://www.mynewsdesk.com/en/blog/5-ways-to-transcribe-audio-to-text/>
104. ГОСТ ISO/IEC 2382-37-2016 Информационные технологии: Словарь, часть 37 – Биометрия
105. ГОСТ Р 58668.11-2019 Информационные технологии: Биометрия - Форматы обмена биометрическими данными, часть 11 – Данные голоса
106. ГОСТ Р 50840-95 Передача речи по трактам связи: Методы оценки качества, разборчивости и узнаваемости
107. Сайт Московского института коррекционной педагогики https://www.defectologiya.pro/zhurnal/leksicheskiy_zapas_razvitie_po_vozrastam/
108. Официальный сайт компании «Наносемантика» <https://nanosemantics.ai/>
109. Рожин, А.Л., Карпов А.А., Ли И.В. Система автоматического распознавания русской речи SIRIUS / «Искусственный интеллект», №3, 2005

Приложение А.

Пример интерфейса разработанного программного обеспечения для сбора данных при проведении исследования эффективности (Мобильное приложение HearingAid)





УТВЕРЖДАЮ

Руководитель образовательного частного учреждения
дополнительного профессионального образования
«Бауманский компьютерный учебный центр «Специалист»
Рясская Валентина Владимировна
«30» октября 2024 г.

АКТ

о внедрении (использовании) результатов
кандидатской диссертационной работы
Бредихина Александра Юрьевича

Настоящий акт о том, что результаты диссертационной работы Бредихина Александра Юрьевича на тему «РАЗРАБОТКА И ИССЛЕДОВАНИЕ МЕТОДОВ И СРЕДСТВ РАСПОЗНАВАНИЯ ЗВУКОВЫХ СИГНАЛОВ ДЛЯ ДОСТУПА СЛАБОСЛЫШАЩИХ АБОНЕНТОВ К РЕСУРСАМ СЕТЕЙ, СИСТЕМ И УСТРОЙСТВ ТЕЛЕКОММУНИКАЦИЙ»,

представленной на соискание ученой степени кандидата наук, использованы при внедрении в учебный процесс приложения для дистанционного обучения слабослышащих в нашей организации.

Внедренное приложение для дистанционного обучения слабослышащих разработано для использования на ПК под управлением ОС Windows доступно для скачивания по адресу - https://www.specialist.ru/hearing_aid.

Результаты диссертационной работы Бредихина Александра Юрьевича позволили соблюсти принцип «цифрового равенства» для слабослышащих студентов, расширить круг потенциальных учащихся, повысить качество предоставляемых услуг и уровень подготовки студентов.

Руководитель образовательного
частного учреждения
дополнительного профессионального
образования «Бауманский
компьютерный учебный центр
«Специалист»

Рясская В.В.



Algorithm electronics OÜ

📍 Harju maakond, Tallinn, Lasnamäe linnaosa, Pallasti tn

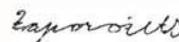
✉ support@algorithm-electronics.com

October 31, 2024

IMPLEMENTATION REPORT

On the Application of Results from the PhD Research of Alexander Y. Bredikhin

This report certifies that the results of Alexander Y. Bredikhin's dissertation research on the topic "Development and Investigation of Methods and Tools for Sound Signal Recognition to Provide Access for Hearing-Impaired Users to Telecommunication Networks, Systems, and Devices," submitted in pursuit of a PhD degree, have been utilized in the implementation of a mobile application for remote auditory training of hearing-impaired users within our organization. The implemented hearing correction application was designed for use on Android smartphones. The results of Alexander Y. Bredikhin's research have enabled our company to offer high-quality and accessible hearing support services to clients across various countries.



CEO, Algorithm Electronics LLC
Aleksandr Zaporozhets

ПЕРЕВОД С АНГЛИЙСКОГО НА РУССКИЙ ЯЗЫК

31 октября 2024 года

ОТЧЕТ О ВНЕДРЕНИИ

О применении результатов диссертационного исследования Александра Ю. Бредихина

Настоящим подтверждается, что результаты диссертационного исследования Александра Ю. Бредихина на тему «Разработка и исследование методов и средств распознавания звуковых сигналов для обеспечения доступа пользователей с нарушениями слуха к телекоммуникационным сетям, системам и устройствам», выполненного в рамках получения ученой степени кандидата наук, были использованы при внедрении мобильного приложения для удаленной аудиотренировки пользователей с нарушениями слуха в нашей организации.

Внедренное приложение для коррекции слуха разработано для использования на смартфонах с операционной системой Android. Результаты исследования Александра Ю. Бредихина позволили нашей компании предоставить высококачественные и доступные услуги поддержки слуха клиентам в различных странах.

Генеральный директор
ООО «Алгоритм Электроникс»
Александр Запорожец

November 11th, 2024



TO WHOM IT MAY CONCERN

CERTIFICATE OF IMPLEMENTATION

For the Use of Research Results from the PhD Dissertation of Alexander Yurievich Bredikhin

This document confirms that the research findings of Alexander Yurievich Bredikhin's PhD dissertation, titled "Development and Study of Methods and Tools for Sound Signal Recognition to Enable Access for Hearing-Impaired Users to Network, System, and Telecommunication Device Resources," have been implemented in the development of a distance learning application for hearing-impaired community in Sri Lanka.

The hearing correction application, designed for use on iOS and Android smartphones, is available at <https://www.dialog.lk/petrallex-app>.

The implementation of Mr. Bredikhin's research has contributed significantly to supporting the principle of "digital equality" for hearing-impaired community in Sri Lanka and users of Dialog Axiata PLC, broadening the range of potential subscribers, and enhancing the quality of services provided.

Yours Faithfully,

Tharusha Wanigasekera
Manager – Digital Innovation and Inclusion

Dialog Axiata PLC (PQ 38)
475, Union Place, Colombo 02, Sri Lanka. Tel +94 (0) 77 7 678 700 www.dialog.lk
an axiata company

Confidential

ПЕРЕВОД С АНГЛИЙСКОГО НА РУССКИЙ ЯЗЫК

11 ноября 2024 г.

Кому это может быть адресовано

СЕРТИФИКАТ О ВНЕДРЕНИИ

на использование результатов исследований из диссертации
Александра Юрьевича Бредихина

Настоящим подтверждается, что результаты исследований, изложенные в диссертации Александра Юрьевича Бредихина на тему «Разработка и исследование методов и средств распознавания звуковых сигналов для обеспечения доступа пользователей с нарушением слуха к ресурсам сетевых, системных и телекоммуникационных устройств», были внедрены в разработку приложения дистанционного обучения для сообщества с нарушением слуха в Шри-Ланке.

Приложение для коррекции слуха, предназначенное для использования на смартфонах под управлением iOS и Android, доступно по адресу: <https://www.dialog.lk/petrallex-app>.

Внедрение результатов исследования г-на Бредихина существенно поспособствовало поддержке принципа «цифрового равенства» для сообщества с нарушением слуха в Шри-Ланке и пользователей компании Dialog Axiata PLC, расширению круга потенциальных абонентов и повышению качества предоставляемых услуг.

С уважением,

Таруша Ванигасекера

Менеджер по цифровым инновациям и инклюзии