

Учреждение образования
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

УДК 621.391; 519.72

ВАШКЕВИЧ
Максим Иосифович

**МЕТОДЫ ЧАСТОТНО-ВРЕМЕННОГО АНАЛИЗА И ОБРАБОТКИ
ЗВУКОВЫХ СИГНАЛОВ НА ОСНОВЕ БАНКОВ ФИЛЬТРОВ
С ФАЗОВЫМ ПРЕОБРАЗОВАНИЕМ**

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
доктора технических наук

по специальности 05.13.17 – Теоретические основы информатики

Минск 2022

Работа выполнена в учреждении образования «Белорусский государственный университет информатики и радиоэлектроники».

Научный консультант **Азаров Илья Сергеевич**, доктор технических наук, доцент, заведующий кафедрой электронных вычислительных средств учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»

Официальные оппоненты **Старовойтов Валерий Васильевич**, доктор технических наук, профессор, главный научный сотрудник лаборатории идентификации систем государственного научного учреждения «Объединенный институт проблем информатики Национальной академии наук Беларуси»

Головко Владимир Адамович, доктор технических наук, профессор, заведующий кафедрой интеллектуальных информационных технологий учреждения образования «Брестский государственный технический университет»

Мурашко Игорь Александрович, доктор технических наук, доцент, профессор кафедры информационных технологий учреждения образования «Гомельский государственный технический университет имени П. О. Сухого»

Оппонирующая организация **Белорусский государственный университет**

Защита состоится 30 июня 2022 г. в 14.00 на заседании совета по защите диссертаций Д 02.15.04 при учреждении образования «Белорусский государственный университет информатики и радиоэлектроники» по адресу: 220013, г. Минск, ул. П. Бровки, 6, корп. 1, ауд. 232, тел. 293-89-89, e-mail: dissovet@bsuir.by.

С диссертацией можно ознакомиться в библиотеке учреждения образования «Белорусский государственный университет информатики и радиоэлектроники».

Автореферат разослан «27» мая 2022 г.

И. о. ученого секретаря совета
по защите диссертаций,
доктор технических наук, профессор



М. М. Татур

ВВЕДЕНИЕ

Анализ, обработка и классификация звуковых сигналов лежат в основе большого числа коммерческих приложений современной IT-индустрии. Известно, что принципы обработки звуковой информации, заимствованные из живых распознающих систем, позволяют совершенствовать технические системы. В настоящее время модели слухового восприятия являются неотъемлемой частью систем кодирования звука, распознавания речи, шумоподавления, повышения качества речевого сигнала, идентификации диктора, распознавания эмоционального состояния диктора, объективной оценки качества речевого сигнала, детектирования неврологических заболеваний по речи и пр.

Исследованиями по анализу и обработке звуковых сигналов занимаются многие ученые Республики Беларусь, СНГ и дальнего зарубежья: Б. Ф. Лобанов, В. Н. Сорокин, А. Л. Ронжин, А. А. Карпов, В. К. Уваров, М. Н. Гусев, J. Kates, F. Baumgarte, В. Moore, T. Quatieri, R. Patterson, T. Irino и др.

Несмотря на значительные научные достижения в данной области, многие вопросы моделирования слухового восприятия, необходимые для систем анализа и обработки звуковых сигналов, остаются нерешенными, что придает развитию данного направления важное социально-экономическое значение. В таких задачах, как коррекция слуховых патологий (в слуховых аппаратах и кохлеарных имплантах), шумоподавление, повышение разборчивости и кодирование критически важна возможность реконструкции обработанного звукового сигнала. Необходимо не только произвести анализ звукового сигнала с учетом свойств слуховой системы человека, но также выполнить модификацию полученного представления сигнала, а затем выполнить реконструкцию. Однако, большинство моделей слухового восприятия позволяют выполнять только анализ звукового сигнала без возможности обратного восстановления. Данное фундаментальное ограничение является побудительной причиной к разработке новых методов психоакустически мотивированного анализа и синтеза звуковых сигналов.

При помощи современных мультимедийных мобильных вычислительных платформ (смартфонов и планшетов) сегодня стало возможным практическое использование сложных методов обработки сигналов. Современные смартфоны оснащены необходимыми аппаратными средствами для создания на их основе различных систем (в том числе и медицинского назначения) обработки звуковых сигналов, причем работающих в режиме реального времени.

Таким образом, создание новых методов моделирования слухового восприятия и соответствующих средств частотно-временного анализа и обработки звуковых сигналов является актуальным научным направлением, формирующим основу для создания новых мультимедиа-систем обработки звуковой информации.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Связь работы с научными программами и темами

Диссертационная работа выполнена в соответствии с научно-техническими заданиями и планами работ кафедры «Электронные вычислительные средства» научно-исследовательской лаборатории 3.1 «Мультипроцессорные системы реального времени» учреждения образования «Белорусский государственный университет информатики и радиоэлектроники» и проводилась в соответствии с государственными научными темами, выполненными в рамках бюджетного финансирования Республики Беларусь:

1. Синтез неравнополосных банков фильтров и их быстрое прототипирование на структуры процессоров слуховых аппаратов / Белорусский республиканский ФФИ, Министерство образования Республики Беларусь; рук. д-р техн. наук А. А. Петровский. – Минск, 2011. – № ГР 20114547. – 15.04.2011–31.03.2013 гг.

2. Методы и мобильные вычислительные структуры обработки сигналов для медицинских приложений / Белорусский республиканский ФФИ, Министерство образования Республики Беларусь; рук. д-р техн. наук И. С. Азаров. – Минск, 2014. – № ГР 20143339. – 23.05.2014–31.03.2016 гг.

3. Методы цифровой коррекции звука на основе принципов психоакустики для повышения качества восприятия в условиях противодействия / Министерство образования Республики Беларусь; рук. д-р техн. наук И. С. Азаров. – Минск, 2017. – № ГР 20171121. – 18.04.2017–31.03.2019 гг.

4. Разработка программного обеспечения скринингового аудиологического обследования детей школьного возраста на базе мультимедийного приложения Petralex / Договор о научно-техническом сотрудничестве между УО «Гродненский государственный медицинский университет» и УО «Белорусский государственный университет информатики и радиоэлектроники» № 17-4043. – 01.09.2017–30.05.2018 гг.

Тема диссертационной работы соответствует пункту 5 «Информатика и космические исследования» перечня приоритетных направлений, определенных Постановлением Совета Министров Республики Беларусь № 190 от 12.03.2015 г. «О приоритетных направлениях научных исследований Республики Беларусь на 2016 – 2020 годы».

Цель и задачи исследования

Целью диссертации является разработка методов частотно-временного анализа и обработки звуковых сигналов, учитывающих частотную и временную разрешающую способность слуха, для решения широкого спектра задач обработки звуковой информации, включающих повышение разборчивости речи, коррекцию слухового восприятия, выявление заболеваний на основе анализа голосового сигнала и пр.

Поставленная цель предполагает решение ряда задач:

1. Разработать методологические подходы к построению моделей слухового восприятия, учитывающие частотные и временные характеристики слуха человека, а также создать алгоритмическую основу для эффективной реализации разработанных подходов в прикладных задачах цифровой обработки сигналов, в том числе и на мобильных вычислительных платформах.

2. Разработать методы обработки сигнала, позволяющие повышать разборчивость речи при нарушениях слухового восприятия, реализуемые на мобильных вычислительных платформах в реальном масштабе времени. При этом осуществляется разработка методов обработки для задач:

– спектральной коррекции звукового сигнала с малой алгоритмической задержкой;

– подавления акустической обратной связи с устойчивостью к резким изменениям параметров пути распространения акустической обратной связи.

3. Разработать метод повышения качества слухового восприятия в шумовой обстановке на основе психоакустических принципов и учитывающий особенности слуха пользователя.

4. Разработать методы акустического анализа речевого/голосового сигнала в кепстральной и модуляционной области с применением моделей слухового восприятия. При этом осуществляется разработка методов обработки для задач:

– выявления заболеваний гортани на основе акустического анализа голосового сигнала;

– автоматического детектирования речевых/голосовых нарушений у пациентов с боковым амиотрофическим склерозом (БАС).

Объектом исследования являются системы обработки речевой и голосовой информации, в том числе и медицинского назначения. *Предметом исследования* выступают методы анализа сигнала, позволяющие моделировать слуховое восприятие акустической информации человеком, а также подходы к реализации систем коррекции слухового восприятия и диагностики заболеваний по голосу на основе мобильных вычислительных платформ.

Научная новизна

1. Предложен новый подход к построению моделей слухового восприятия, основанный на использовании банков фильтров с фазовым преобразованием, что позволило, используя сходство частотно-временного покрытия неравнополосного банка фильтров с моделями на основе слуховых гамматон-фильтров, разработать эффективный метод неравнополосной частотной декомпозиции и обработки сигнала с возможностью обратной реконструкции. Получаемая в результате модель обладает большей прикладной ценностью по сравнению с моделью слухового восприятия, построенной на основе гамматон-фильтров.

2. Разработан оригинальный метод коррекции слухового восприятия, основанный на частотно-временной декомпозиции, использующей модифицированное дискретное косинусное преобразование с деформированной частотной шкалой, обладающий малой алгоритмической задержкой, позволяющий выполнять коррекцию слуха на мобильной вычислительной платформе с эффективностью, близкой к современным цифровым слуховыми аппаратами.

3. Предложен метод подавления акустической обратной связи с высокой устойчивостью к резким изменениям характеристик пути распространения обратной связи, основанный на модифицированном правиле спектрального вычитания, что позволяет строить эффективные системы коррекции слухового восприятия на основе мобильных платформ.

4. Разработан метод обработки звуковых сигналов (речевых и музыкальных), основанный на использовании неравнополосного банка фильтров с фазовым преобразованием и субканального нелинейного усиления, зависящего от интенсивности и характера зашумления, что позволяет воспроизводить звук с повышенной разборчивостью в условиях фонового шума.

5. Разработан алгоритм извлечения информационно значимого параметра голосового сигнала, характеризующих степень патологичности вибрато, основанный на количественной оценке частотных компонент вибрато и учитывающий особенности изменения голоса у пациентов с БАС, позволяющий повысить точность систем детектирования данного заболевания по речи.

6. Предложен метод вычисления барк-частотных кепстральных коэффициентов (БЧКК), основанный на использовании неравнополосного ДПФ-модулированного банка фильтров. Показано, что использование БЧКК в сочетании с пертурбационными параметрами голоса позволяет строить эффективные системы классификации пациентов с неврологическим заболеванием БАС со средней полнотой классификации на уровне 96,7 %.

7. Предложен метод анализа модуляционного спектра звукового сигнала, использующий неравнополосный ДПФ-модулированный банк фильтров с фазовым преобразованием для частотно-временной декомпозиции. Показано, что применение данного метода анализа в системе детектирования патологии голосового аппарата позволяет обеспечивать более высокую точность классификации, по сравнению с использованием традиционного метода анализа модуляционного спектра на основе ДПФ с объединением энергий в субполосах.

Положения, выносимые на защиту

1. Новый подход к построению средств цифровой обработки звука, учитывающий свойства слуховой системы человека, основанный на использовании неравнополосных банков фильтров с фазовым преобразованием, позволяющих выполнять частотно-временную декомпозицию звукового сигнала с последующим вос-

становлением. Частотно-временное покрытие разработанных банков фильтров с фазовым преобразованием имитирует свойства слуховой системы человека аналогично банку гамматон-фильтров. Получаемые в результате средства обработки звука являются вычислительно эффективными и имеют значительно более широкую сферу применения по сравнению с подходом на основе банка гамматон-фильтров, который не обеспечивает восстановления сигнала.

2. Метод коррекции слуховых нарушений с малой алгоритмической задержкой (менее 5 мс) для мобильных вычислительных платформ, основанный на частотно-временной декомпозиции с использованием модифицированного дискретного косинусного преобразования с деформированной частотной осью и суммировании субканальных сигналов, обеспечивающий повышение эффективности коррекции слуха на 2,3 % по сравнению с применением современных слуховых аппаратов.

3. Метод подавления акустической обратной связи (АОС) для систем коррекции слухового восприятия, ориентированный на мобильные платформы, основанный на модифицированном правиле спектрального вычитания и обладающий высоким уровнем устойчивости к резким изменениям характеристик пути распространения обратной связи. Метод позволяет получить сегментное отношение сигнал/шум в среднем на 9,8 дБ выше, чем у широкополосной системы со стандартной техникой подавления АОС на основе адаптивной фильтрации и метода наименьших квадратов.

4. Метод повышения качества восприятия воспроизводимого звукового сигнала в условиях шума, основанный на неравнополосном банке фильтров с фазовым преобразованием и разработанном правиле нелинейного усиления частотно-временных компонент воспроизводимого сигнала, зависящего от спектральных характеристик окружающего шума. В агрессивной шумовой обстановке (ОСШ меньше –20 дБ) метод обеспечивает индекс разборчивости речи (ИРР) более 0,45, а при снижении интенсивности шума (ОСШ более 0 дБ) – хорошую разборчивость (ИРР более 0,75).

5. Алгоритм извлечения из голосового сигнала параметра патологичности вибрато, учитывающий особенности изменения голоса у пациентов с БАС. Использование параметра патологичности вибрато обеспечивает среднюю полноту классификации на уровне 90,3 %, что на 2,5 % выше, чем у классификатора, использующего базовый набор акустических признаков, состоящий из коэффициентов частотной и амплитудной нестабильности.

6. Метод вычисления барк-частотных кепстральных коэффициентов (БЧКК) голосового сигнала, основанный на неравнополосном ДПФ-модулированном банке фильтров, применение которых совместно с пертурбационными параметрами голоса позволяет строить эффективные системы классификации пациентов с

БАС-патологией со средней полнотой классификации на уровне 96,7 %, что на 2,6 % выше, чем у систем использующих мел-частотные кепстральные коэффициенты.

7. Оригинальный метод анализа модуляционного спектра звукового сигнала на основе неравнополосного банка фильтров с фазовым преобразованием, применение которого повышает эффективность классификации патологии голосового аппарата по сравнению с использованием общепринятого метода анализа на основе ДПФ с объединением энергий в субполосах.

Личный вклад соискателя ученой степени

Основные результаты и положения, выносимые на защиту, получены соискателем лично. Начальная постановка цели и задач исследования выполнена научным консультантом д.т.н. профессором Петровским А. А. Профессор Петровский А.А. участвовал в обсуждении результатов исследований, полученных до 2019 г. В связи со смертью А. А. Петровского в 2019 г. научным консультантом назначен д.т.н. доцент Азаров И. С. Вклад Азарова И. С. заключался в корректировке некоторых целей и задач диссертационного исследования, а также обсуждении результатов, полученных соискателем. В представленных публикациях с соавторами вклад соискателя определяется рамками излагаемых в диссертации результатов.

Апробация диссертации и информация об использовании ее результатов

Результаты, полученные в ходе выполнения исследований, докладывались и обсуждались на 15-й, 16-й, 20-й и 23-й международных конференциях «Цифровая обработка сигналов и ее применение» (Москва, Россия, 2013, 2014, 2018, 2021); 14th International Conference Pattern Recognition and Information Processing (Minsk, Belarus, 2019); Международной научно-технической конференции, приуроченной к 50-летию МРТИ–БГУИР (Минск, Беларусь, 2014); International Conference on Audio, Language and Image Processing (ICALIP'2014) (Shanghai, China, 2014); Международной научно-технической конференции «Медэлектроника» (Минск, Беларусь, 2014, 2018); IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'2016) (Shanghai, China, 2016); IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'2017) (Louisiana, USA, 2017); International Conference Signal Processing Algorithms, Architectures, and Applications (SPA) (Poznan, Poland, 2017–2019); 6-й Международной конференции «BIG DATA и анализ высокого уровня» (Минск, Беларусь, 2020); 10-й Международной научной конференции «Информационные технологии и системы» (Минск, Беларусь, 2020); 23rd international Conference on Speech and Computer (SPECOM 2021) (St. Petersburg, Russia, 2021).

Опубликование результатов диссертации

По тематике диссертации опубликованы 54 печатные работы, в том числе 1 монография, 34 статьи в рецензируемых научных журналах, 19 статей в сборниках материалов научных конференций. Суммарный объем публикаций составляет 39,4 авторского листа. Результаты работы включены в 3 отчета по НИР.

Структура и объем диссертации

Диссертация включает в себя введение, общую характеристику работы, семь глав, содержащие аналитический обзор литературы и результаты оригинальных исследований, заключение с рекомендациями по практическому использованию полученных результатов и библиографический список. Полный объем диссертации составляет 269 страницы. Диссертация содержит 127 рисунков на 39 страницах и 24 таблицы на 10 страницах. Библиографический список включает 239 наименований и занимает 20 страниц.

ОСНОВНАЯ ЧАСТЬ

Первая глава посвящена рассмотрению существующих методов частотно-временного анализа звуковых сигналов, основанных на моделировании свойств слухового восприятия человека. Модели слухового восприятия лежат в основе систем кодирования звука, распознавания речи, шумоподавления, повышения качества речевого сигнала, идентификации диктора, распознавания эмоционального состояния диктора, объективной оценки качества речевого сигнала, детектирования неврологических заболеваний по речевым образцам и пр. Обобщенный вид модели слуха представлен на рисунке 1.

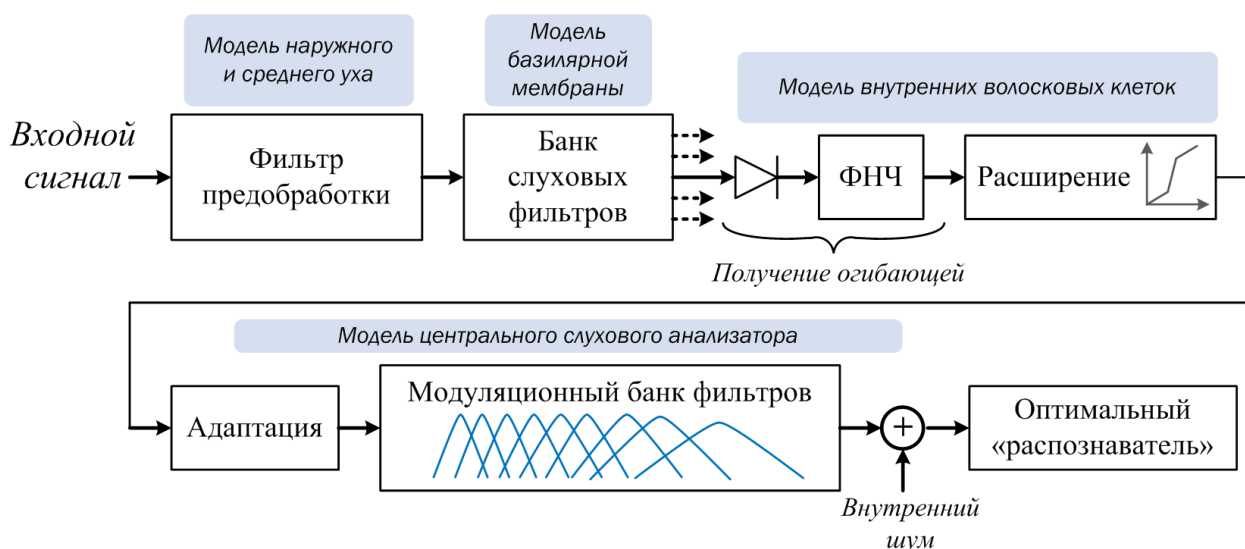


Рисунок 1. – Обобщенная модель слухового восприятия

Ключевая роль в обработке звука в слуховой системе человека принадлежит улитке внутреннего уха. Базилярная мембрана улитки реагирует на акустические колебания таким образом, что каждой частоте соответствует участок на

базиллярной мембране. С технической точки зрения можно считать, что мембрана осуществляет частотно-временную декомпозицию сигнала. Данный процесс моделируется (см. рисунок 1) при помощи банка слуховых фильтров.

В качестве альтернативы данному подходу широко применяются методы частотно-временной декомпозиции, основанные на закономерностях, выявленных в результате психоакустических экспериментов. В первой главе выполнен обзор методов частотно-временного анализа согласно данной категоризации: методы, моделирующие отклик базиллярной мембраны (т.е. слуховые фильтры), и психоакустически мотивированные методы [54]. Показано, что в основе моделей слуховых фильтров лежат данные физиологических экспериментов. Проанализированы наиболее распространенные модели слуховых фильтров: *roex*-фильтры, гамматон-фильтры, гаммачирп-фильтры, кохлеарная модель Кейтса. В психоакустически мотивированных методах анализа ставится задача получения частотной декомпозиции, соответствующая критическим полосам восприятия, либо аппроксимирующей какую-либо неравномерную психоакустическую шкалу (например, барки, мелы, ERB и пр.). Исследована возможность получения неравнополосной частотной декомпозиции сигнала при помощи дискретного преобразования Фурье (ДПФ) с объединением энергий в субканалах, а также банка фильтров на основе вейвлет-преобразования.

Проведенный анализ показал, что особенностью методов, моделирующих отклик базиллярной мембраны (*roex*-фильтры, гамматон-фильтры, гаммачирп-фильтры, модель Кейтса) является то, что они хорошо согласованы с физиологическими данными о работе слуховой системы. Однако для данной категории методов отсутствуют процедуры восстановления сигнала. Это существенно ограничивает область применения данных методов задачами распознавания образов (например, распознавание/классификация речи и идентификация диктора).

Также показано, что практическая реализация методов, моделирующих отклик базиллярной мембраны, требует значительных вычислительных затрат, что также ограничивает область их применения.

Особенностью психоакустически мотивированных методов анализа звукового сигнала является наличие эффективных алгоритмов реализации. Показано, что в методе ДПФ с объединением энергии в субполосах временное разрешение определяется размером преобразования (т. е. длиной ДПФ) и, таким образом, является постоянным для всех анализируемых частотных полос. Для высокочастотных полос это приводит к значительно более низкому временному разрешению по сравнению с соответствующими слуховыми фильтрами. Использование банков фильтров на основе вейвлет-преобразования позволяет получить частотное разрешение с постоянной добротностью. Эти банки фильтров также имеют неравномерное временное разрешение, что отражает свойство реальной слуховой систе-

мы. Недостатком банков фильтров на основе вейвлет-преобразования является то, что производимая ими частотно-временная декомпозиция сигнала напрямую зависит от древовидной структуры банка фильтров. Также ширина критических полос недостаточно точно моделируется при помощи таких банков фильтров. Таким образом, требуется разработка методов частотно-временного анализа сигнала, которые учитывают как временное, так и частотное разрешение слуховой системы человека и не имеют привязки к структуре схемы реализации.

На основании проведенного анализа делается вывод, что для улучшения качества работы систем, основанных на использовании моделей слухового восприятия, требуется разработка метода частотно-временного анализа звукового сигнала, который обеспечивает хорошую аппроксимацию частотно-временного разрешения слуха; имеет эффективный алгоритм реализации; обеспечивает восстановление сигнала (как в случае банка фильтров на основе вейвлет-преобразования); обеспечивает возможность гибкой настройки частотно-временного разрешения без жесткой привязки к структурной реализации (в отличие от банка фильтров на основе вейвлет-преобразования).

Вторая глава посвящена разработке методов синтеза неравнополосных банков фильтров на основе фазового преобразования и принципа модуляции [1, 2, 22, 36]. Неравнополосные банки фильтров, основанные на фазовом преобразовании, сочетают в себе преимущества кратковременного преобразования Фурье (КВПФ) [22, 33], а также вейвлет-преобразования [30]. Как и при практической реализации КВПФ, в них можно использовать эффективные алгоритмы быстрого преобразования Фурье (БПФ) и при этом достигать частотного разрешения с постоянной добротностью. Их важной особенностью является гибкость – есть всего один параметр, отвечающий за степень деформации частотной оси. Изменяя данный параметр, можно получать различные конфигурации покрытия частотно-временной плоскости. Этим они отличаются от банков фильтров на основе вейвлет-преобразования, где частотно-временное покрытие определяется структурой банка фильтров.

Рассмотрены два типа банков фильтров, основанных на принципе модуляции: ДПФ-модулированный банк фильтров и косинусно-модулированный банк фильтров (КМБФ). M -канальный ДПФ-модулированный банк фильтров образуется путем модуляции импульсной характеристики $h(n)$ КИХ-фильтра-прототипа нижних частот с частотой среза $\frac{\pi}{M}$:

$$h_k(n) = h(n)W_M^{-nk}, \quad (1)$$

где $W_M = \exp(-j2\pi/M)$; $k = 0, 1 \dots M - 1$ – номер канала; $n = 0, 1 \dots N - 1$, $N = mM$ – порядок фильтра-прототипа; $m \in \mathbb{N}$.

Выражение (1) описывает банк фильтров анализа, банк фильтров синтеза задается как $f_k(n) = h_k^*(N - 1 - n)$. ДПФ-модулированный банк фильтров может

быть эффективно реализован в виде полифазной структуры, описываемой выражением [6, 36]

$$\mathbf{H}(z) = \text{IDFT}_M \cdot \underbrace{[\mathbf{I}_M \quad \mathbf{I}_M \quad \dots]}_{m \text{ повторов}} \cdot \text{diag}(1, z^{-1}, \dots, z^{N-1}) \cdot \mathbf{h}, \quad (2)$$

где $\mathbf{H}(z) = [H_0(z) \ H_1(z) \ \dots \ H_{M-1}(z)]^T$ – вектор z -образов фильтров анализа; IDFT_M – матрица $M \times M$ обратного ДПФ; \mathbf{I}_M – единичная матрица размером $M \times M$; $\mathbf{h} = [h(0) \ h(1) \ \dots \ h(N-1)]^T$ – вектор коэффициентов фильтра-прототипа.

Выходами ДПФ-модулированного банка фильтров являются комплексные субполосные сигналы, что не всегда соответствует решаемым задачам [9, 13, 22]. Для получения банка фильтров с вещественными субполосными сигналами используется косинусная модуляция. M -канальный КМБФ анализа и синтеза описывается выражениями

$$h_k(n) = 2h(n) \cos\left(\frac{(2k+1)\pi}{2M}\left(n - \frac{N-1}{2}\right) + \theta_k\right), \quad (3)$$

$$f_k(n) = 2h(n) \cos\left(\frac{(2k+1)\pi}{2M}\left(n - \frac{N-1}{2}\right) - \theta_k\right), \quad (4)$$

где $h(n)$ – коэффициенты фильтра-прототипа нижних частот с частотой среза $\frac{\pi}{2M}$; $N = 2mM$ – порядок фильтра-прототипа; $\theta_k = (-1)^k(\pi/4)$.

КМБФ также имеет эффективную полифазную реализацию, описываемую выражением

$$\mathbf{H}(z) = \mathbf{C} \cdot \underbrace{[\mathbf{I}_{2M} \quad -\mathbf{I}_{2M} \quad \mathbf{I}_{2M} \quad -\mathbf{I}_{2M} \quad \dots]}_{m \text{ повторов}} \cdot \text{diag}(1, z^{-1}, \dots, z^{N-1}) \cdot \mathbf{h}, \quad (5)$$

где $\mathbf{C} = [c_{k,n}]_{0 \leq k < M, 0 \leq n < 2M}$, $c_{k,n} = 2 \cos\left(\frac{\pi(k+0,5)}{M}\left(n - \frac{N-1}{2}\right) + \theta_k\right)$.

Эффективность полифазной реализации КМБФ (5) заключается в наличии быстрого алгоритма умножения на матрицу косинусной модуляции \mathbf{C} . Под быстрым алгоритмом понимается представление матрицы \mathbf{C} в виде произведения разреженных матриц:

$$\mathbf{C} = \begin{cases} \sqrt{M}(-1)^{m_1} \text{DCT-4}_M[(\mathbf{I}_M - \mathbf{J}_M) \quad -(\mathbf{I}_M + \mathbf{J}_M)], \\ \text{для четного } m = 2m_1, \\ \sqrt{M}(-1)^{m_1} \text{DCT-4}_M[(\mathbf{I}_M + \mathbf{J}_M) \quad (\mathbf{I}_M - \mathbf{J}_M)], \\ \text{для нечетного } m = 2m_1 + 1, \end{cases} \quad (6)$$

где DCT-4_M – матрица дискретного косинусного преобразования (ДКП) четвер-

того типа; \mathbf{J}_M – матрица $M \times M$, содержащая единицы на побочной диагонали:

$$\text{DCT-4}_M = \left[\sqrt{\frac{2}{M}} \cos \left(\frac{\pi(k + \frac{1}{2})(\ell + \frac{1}{2})}{M} \right) \right]_{0 \leq k, \ell < M}, \quad \mathbf{J}_M = \begin{bmatrix} & & & 1 \\ & & & \\ & & \cdot & \\ & & & \\ 1 & & & \end{bmatrix}.$$

Выражение (6) позволяет свести задачу умножения на матрицу модуляции \mathbf{C} к задаче вычисления ДКП-4, для которого имеются быстрые алгоритмы.

Далее, в главе 2 рассматриваются математические основы фазового преобразования, показывается его связь с разложением сигнала в базисе дискретных последовательностей Лагерра $\lambda^{(k)}(n)$, z -преобразование которых имеет вид

$$\Lambda^{(k)}(z) = \sqrt{1 - \alpha^2} \frac{(z^{-1} - \alpha)^k}{(1 - \alpha z^{-1})^{k+1}}, \quad (7)$$

где $|\alpha| < 1$ – регулируемый параметр.

Из (7) следует, что z -преобразование последовательностей Лагерра удовлетворяет рекуррентному соотношению

$$\Lambda^{(k+1)}(z) = A(z)\Lambda^{(k)}(z) = A^k(z)\Lambda^{(0)}(z), \quad (8)$$

где

$$A(z) = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}} \quad (9)$$

передаточная функция всепропускающего звена первого порядка. Частотная характеристика фазового звена равна $A(e^{j\omega}) = e^{-j\theta(\omega)}$, где

$$\theta(\omega) = \omega + \arctg((\alpha \sin \omega)/(1 - \alpha \cos \omega)). \quad (10)$$

Благодаря (8) перевод сигнала в базис дискретных последовательностей Лагерра может быть осуществлен пропусканием сигнала $x(n)$ через цепочку фазовых звеньев, в результате сигнал представляется коэффициентами Лагерра u_k :

$$X(z) = \sum_{k=0}^{\infty} u_k \Lambda^{(k)}(z). \quad (11)$$

Исходя из (11) и (8), а также учитывая частотную характеристику фазового звена, можно записать выражение для Фурье-образа сигнала $x(n)$:

$$X(e^{j\omega}) = \Lambda^{(0)}(e^{j\omega}) \sum_{k=0}^{\infty} u_k A^k(e^{j\omega}) = \Lambda^{(0)}(e^{j\omega}) U(e^{j\theta(\omega)}), \quad (12)$$

где

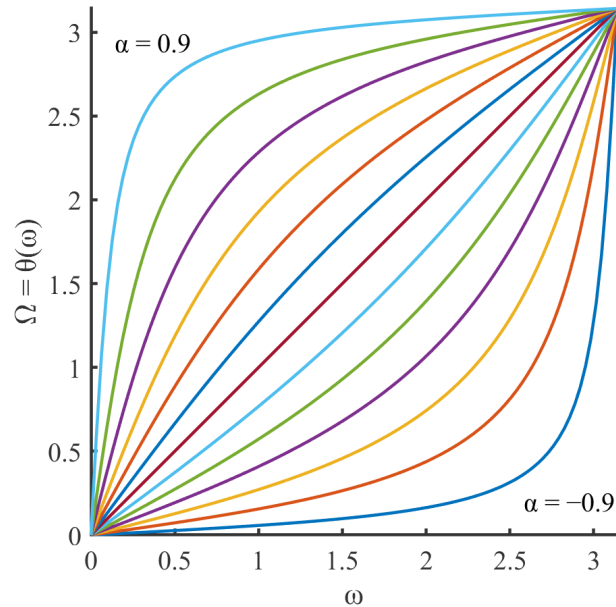


Рисунок 2. – Семейство кривых, описывающих деформацию частотной оси

$$U(e^{j\theta(\omega)}) = \sum_{k=0}^{\infty} u_k e^{-jk\theta(\omega)}, \quad (13)$$

а $\theta(\omega)$ задается выражением (10).

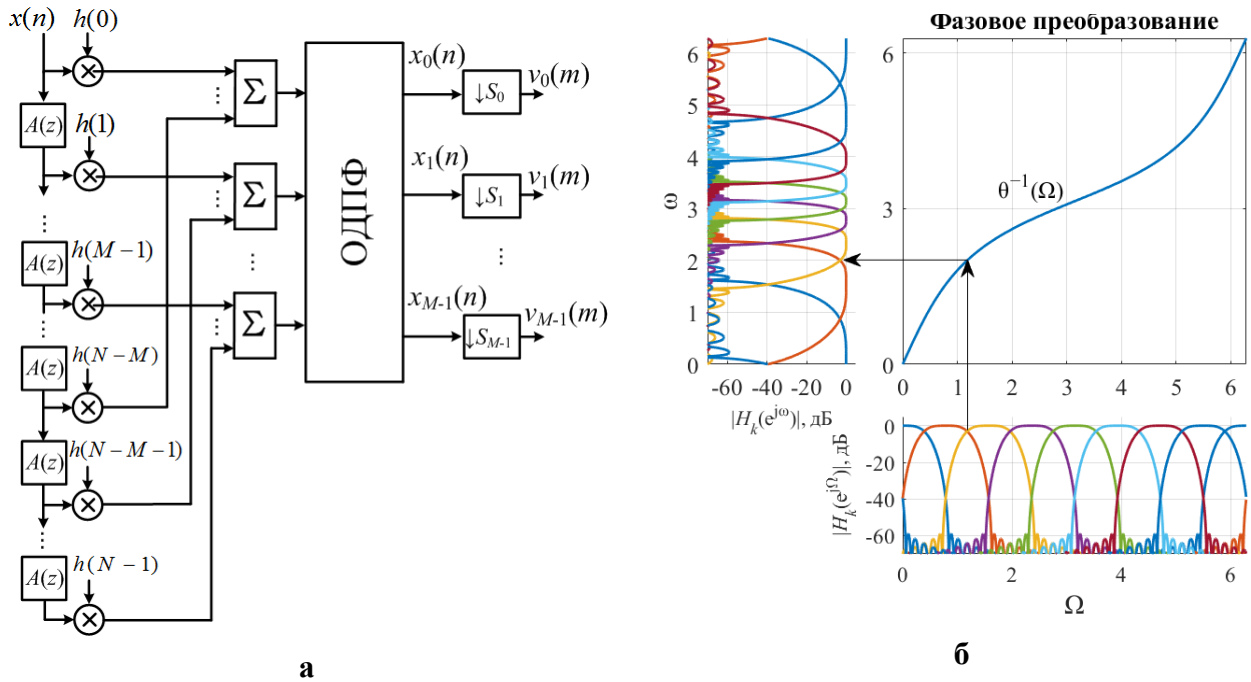
Выражение (12) показывает, что Фурье-образ последовательности коэффициентов Лагерра $U(e^{j\Omega})$ является частотной деформированной версией взвешенного частотного спектра $X(e^{j\omega})$. Сама деформация частотной оси $\Omega = \theta(\omega)$ управляется параметром α фазового звена (9). В z -области это отображение описывается заменой

$$z^{-1} \rightarrow A(z), \quad (14)$$

которое является уникальным однозначным отображением единичной окружности на себя, чей модуль равен единице. На рисунке 2 показано семейство кривых, описывающих деформацию частотной оси, происходящую вследствие отображения (14).

Фазовое преобразование (14) применяется к полифазной структуре банков фильтров (2) и (5) для получения их неравнополосных версий. На рисунке 3 показана структура неравнополосного ДПФ-модулированного банка фильтров, а также деформация оси частот, происходящая в процессе фазового преобразования.

Для изучения различных вариантов банков фильтров в диссертации предложено использовать понятие *частотно-временного покрытия*. Частотно-временное покрытие визуализировалось при помощи прямоугольников Гейзенберга, которые характеризуют временную и частотную протяженность (локализацию) характеристик банка фильтров. Частотная протяженность вычислялась как



а – полифазная структура; б – процесс частотной деформации

Рисунок 3. – Неравнополосный ДПФ-модулированный банк фильтров

$$\sigma_{\omega}^2 = \int_{-\pi}^{\pi} (\omega - \omega_0)^2 |H_k(e^{j\omega})|^2 d\omega, \quad (15)$$

где ω_0 – центральная частота полосы пропускания фильтра $H_k(e^{j\omega})$.

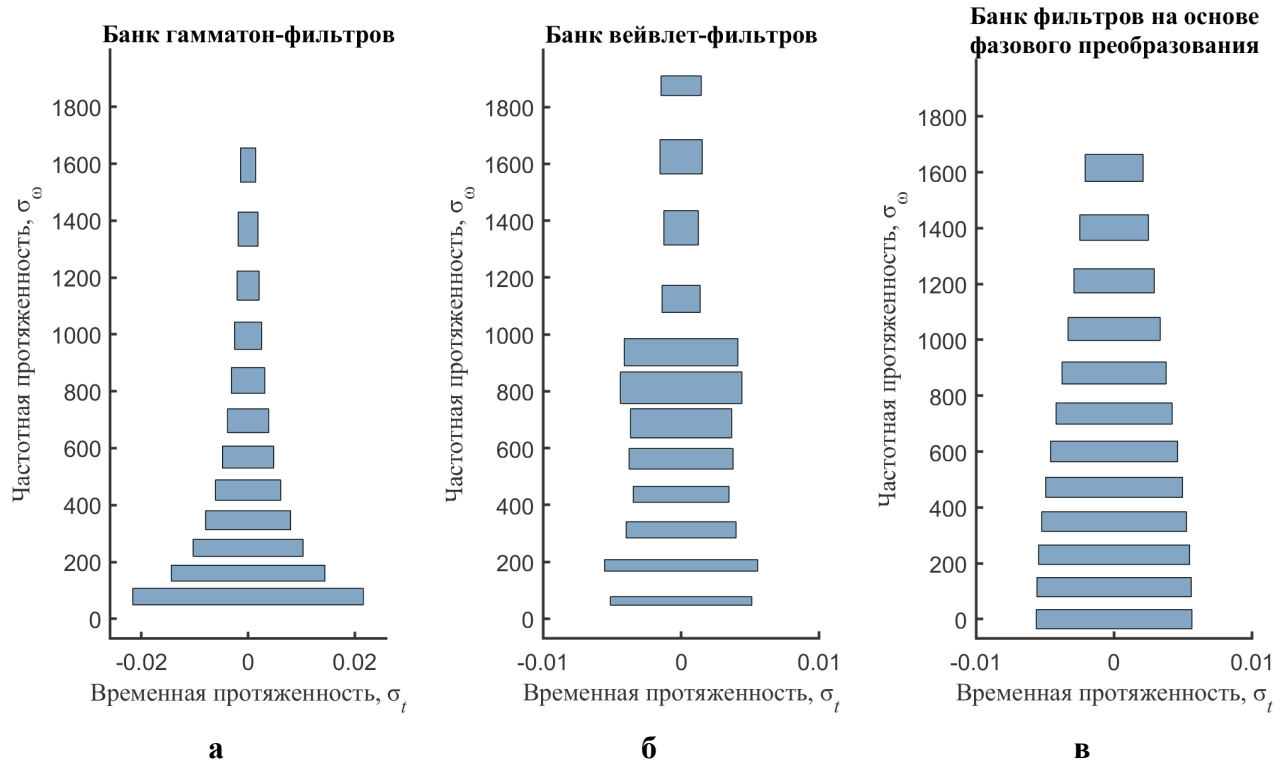
Временная протяженность определялась как

$$\sigma_t^2 = \int_{-\infty}^{\infty} (t - \tau_0)^2 |h_k(t)|^2 dt, \quad (16)$$

где τ_0 – групповая задержка k -го фильтра на частоте ω_{0k} .

На рисунке 4 представлены частотно-временные покрытия трех банков фильтров: 1) банк гамматон-фильтров; 2) банк фильтров на основе вейвлет-преобразования; 3) неравнополосный ДПФ-модулированный банк фильтров.

Из приведенных на рисунке 4 графиков можно видеть, что частотная протяженность гамматон-фильтров увеличивается с увеличением центральной частоты фильтра, а временная протяженность уменьшается с ростом центральной частоты. Это наблюдение согласуется с тем фактом, что частотное разрешение человеческого уха лучше в диапазоне низких частот, чем в диапазоне высоких частот. Частотно-временная локализация банка вейвлет-фильтров имеет нерегулярный вид. Фильтры низких частот имеют лучшее разрешение по частоте, а фильтры высоких частот – лучшее разрешение по времени. Однако вейвлет-фильтры не позволяют моделировать плавное изменение частотно-временной локализации энергии, присущее реальным слуховым фильтрам. В отличие от вейвлет-фильтров ДПФ-модулированный банк фильтров на основе фазового преобразова-



а – банк гамматон-фильтров; б – банк вейвлет-фильтров; в – неравнополосный ДПФ-модулированный банк фильтров на основе фазового преобразования

Рисунок 4. – Частотно-временное покрытие

ния обеспечивает плавное изменение частотно-временной локализации энергии. Прямоугольники Гейзенберга, полученные для этого банка фильтров, очень похожи на те, что получены для гамматон-фильтров. Таким образом, банк фильтров на основе фазового преобразования можно рассматривать как альтернативу банку гамматон-фильтров, поскольку он имеет эффективную схему реализации.

В **третьей главе** рассмотрены вопросы практического использования разработанных неравнополосных банков фильтров на основе фазового преобразования. Показано, как выполняется выбор параметра фазового преобразования в зависимости от вида психоакустической шкалы, которую должен аппроксимировать банк фильтров. Уделено внимание выбору коэффициентов децимации/интерполяции в системе анализ/синтез на основе банка фильтров. Данный вопрос имеет большое значение, поскольку, с одной стороны, увеличение коэффициентов децимации/интерполяции ведет к увеличению эффективности субполосной обработки за счет снижения скорости потока данных. С другой стороны, неправильный выбор коэффициентов децимации может привести к появлению значительных компонент наложения спектра в сигнале. В результате применения фазового преобразования ФЧХ системы анализ/синтез приобретает нелинейный вид, что не всегда желательно. Для устранения данного эффекта в третьей главе рассмотрен вопрос выравнивания групповой задержки системы анализ/синтез.

Получение недорогой, высокопроизводительной дискретной системы с малым энергопотреблением, как правило, основывается на реализации ее с использованием арифметики с фиксированной запятой. В случае реализации дискретной системы на основе фазового преобразования критически важным элементом является цепочка фазовых звеньев (ЦФЗ), которая представляет из себя каскад БИХ-фильтров первого порядка. В зависимости от конкретного приложения ЦФЗ может содержать от нескольких десятков до нескольких сотен элементов [1, 30, 43], поэтому пятый раздел третьей главы посвящен вопросу реализации ЦФЗ. Исследовались несколько подходов к реализации ЦФЗ с использованием арифметики с фиксированной запятой с применением различных схем квантования [32].

Цепочка фазовых звеньев, анализируемая в диссертации, состоит из каскада БИХ-фильтров первого порядка с передаточной функцией (9). Передаточная функция (9) соответствует разностному уравнению

$$y(n) = x(n - 1) + \alpha(y(n - 1) - x(n)). \quad (17)$$

В соответствии с (17) блок-схему ЦФЗ можно представить в виде, показанном на рисунке 5.

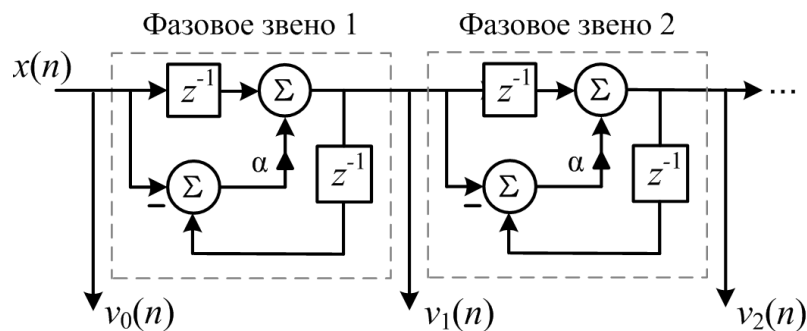


Рисунок 5. – Блок-схема цепочки фазовых звеньев

На рисунке через $x(n)$ обозначен входной сигнал, а через $v_0(n), v_1(n), \dots$ обозначены выходы цепочки фазовых звеньев. При реализации ЦФЗ с использованием B -разрядной арифметики с фиксированной запятой для предотвращения неограниченного роста разрядности промежуточных данных необходимо выполнять квантование результатов арифметических операций. Анализировались следующие варианты квантования [32]: усечение, детерминированное округление, вероятностное округление, использование схемы формирования спектра ошибки.

Для исследования рассмотренных подходов к квантованию результатов арифметических операций в ЦФЗ была выполнена программная реализация данных методов на языке C++. В частности, моделировалась цепочка, состоящая из 180 звеньев, коэффициент $\alpha = 0,4092$, что соответствует фазовому преобразованию, аппроксимирующему шкалу барков для частоты дискретизации 8 кГц. В качестве входных тестовых сигналов рассматривались: 1) дельта-импульс;

2) единичный скачок; 3) синусоидальный сигнал; 4) последовательность дельта-импульсов. В процессе эксперимента рассчитывался сигнал ошибки $e(n)$, который формировался как разность между выходами опорной реализации ЦФЗ и реализации ЦФЗ с фиксированной запятой. Получающийся сигнал ошибки трактовался как шум квантования. Моделирование ЦФЗ производилось для разрядности данных $B = \{10, 12, 14, 16\}$ бит. Сигналы шума квантования $e(n)$, возникающие при использовании различной разрядности данных, сравнивались, опираясь на их среднеквадратичное значение (MSE), соотнесенное с шагом квантования q :

$$\text{MSE} = 10 \log_{10} \left(\sqrt{1/N \sum_{n=0}^{N-1} e^2(n)/q} \right). \quad (18)$$

На рисунке 6 представлены сигналы шумов квантования, возникающих в ЦФЗ при использовании указанных выше методов квантования.

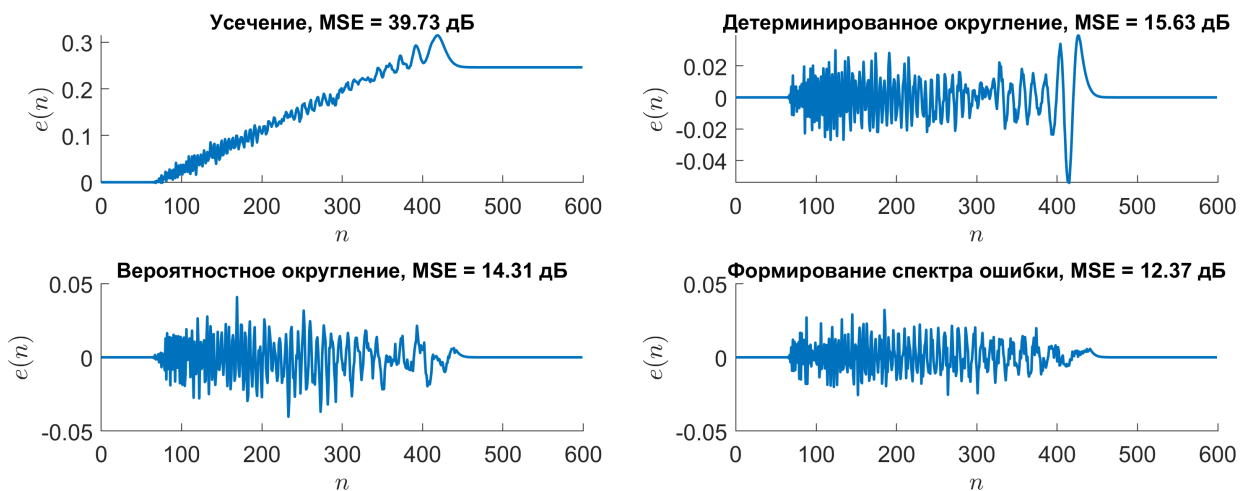


Рисунок 6. – Шум квантования, возникающий в ЦФЗ при подаче на вход дельта-импульса (разрядность данных $B = 10$ бит)

Проведенные эксперименты показали, что шум квантования существенно зависит от вида тестового сигнала. Наибольший уровень шума квантования зафиксирован для синусоидального входного сигнала, а наименьший – при подаче на вход дельта-импульса.

В среднем наименьший уровень шума (для различной разрядности представления данных) дает использование схемы формирования спектра ошибки, а наихудший – квантование с использованием усечения.

С точки зрения практической реализации (аппаратной и программной) наиболее предпочтительным является использованием детерминированного округления как наименее затратного с вычислительной точки зрения. В отличие от схемы формирования спектра ошибки детерминированное округление не требует введения дополнительных элементов памяти и сумматоров. Проведенные эксперименты показали, что при использовании в качестве тестового сигнала

дельта-импульса MSE шума квантования в случае применения детерминированного округления на 1,07 дБ больше, чем MSE шума квантования в случае применения схемы формирования спектра ошибки. Для других тестовых сигналов разница между уровнями шумов квантования меньше 1 дБ.

В четвертой главе описана разработанная система коррекции слуха на основе мобильной вычислительной платформы [1, 8, 16, 17, 19, 38, 39]. Предложена схема обработки сигнала с малой алгоритмической задержкой, позволяющая выполнять коррекцию слуха при нейросенсорной тугоухости и учитывающая особенности вычислительной платформы. Схема позволяет выполнять компенсацию потери слуха путем субполосного частотно-зависимого усиления и компрессии динамического диапазона (КДД) сигнала. Для учета амплитудно-частотных характеристик динамиков гарнитуры аудиометрия выполняется непосредственно при помощи мобильной платформы.

Особенностью предложенной системы коррекции слуха является использование метода суммирования субполосных сигналов на базе модифицированного косинусного преобразования (МДКП) с деформированной частотной осью [1, 43].

МДКП можно описать как банк фильтров следующим выражением:

$$\mathbf{H}(z) = \text{DCT-4}_M \begin{bmatrix} & & -\mathbf{J}_{\frac{M}{2}} & -\mathbf{I}_{\frac{M}{2}} \\ \mathbf{I}_{\frac{M}{2}} & -\mathbf{J}_{\frac{M}{2}} & & \end{bmatrix} \cdot \text{diag}(1, z^{-1}, \dots, z^{N-1}) \cdot \mathbf{h}. \quad (19)$$

Выражение (19) является основой для получения вычислительно эффективной реализации МДКП. Деформация частотной оси выполняется при помощи фазового преобразования (14). На рисунке 7 приведена частотная характеристика неравнополосной версии 22-канального банка фильтров на основе МДКП. Параметр деформации частотной оси α выбирался таким образом, чтобы банк фильтров аппроксимировал психоакустическую шкалу барков.

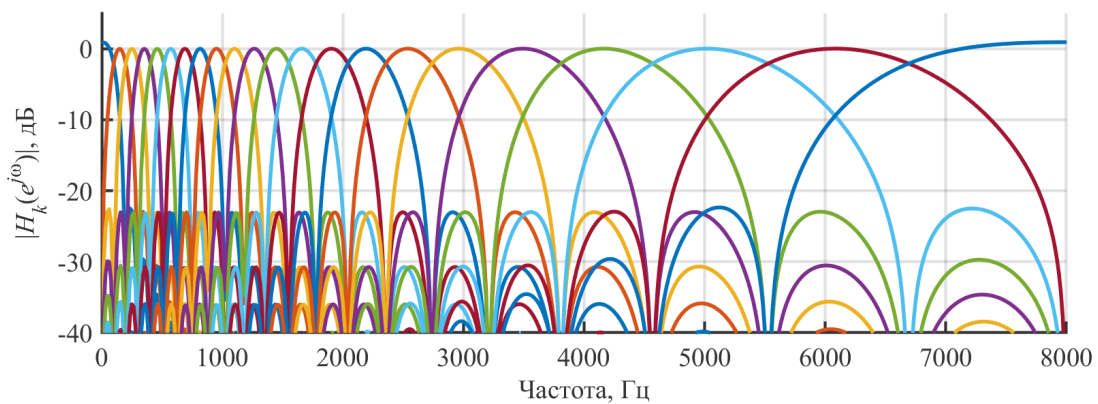


Рисунок 7. – Частотная характеристика МДКП с деформацией частотной оси

На рисунке 8 показана схема обработки сигнала на основе метода суммирования субканальных сигналов с применением МДКП с деформированной ча-

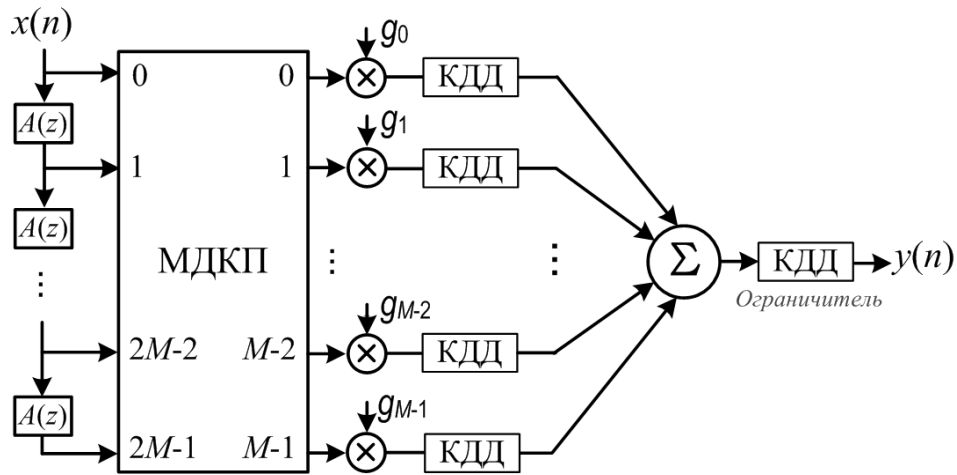


Рисунок 8. – Схема обработки сигнала в системе коррекции слуха на базе мобильной платформы

стотной осью. Коэффициенты g_0, \dots, g_{M-1} используются для формирования заданного частотного профиля. Поскольку сигналы в каналах не децимируются, то синтез выходного сигнала выполняется путем суммирования всех субканальных сигналов. Особенность предлагаемой схемы в том, что она значительно сокращает групповую задержку сигнала, что, как отмечалось, весьма важно для систем коррекции слуха. Так, для частоты дискретизации 16 кГц средняя групповая задержка системы на рисунке 8 составляет порядка 4 мс (рисунок 9). Тот факт, что задержка больше на нижних частотах, чем на верхних, не играет негативной роли, поскольку слуховая система человека на нижних частотах имеет разрешающую способность хуже, чем на верхних.

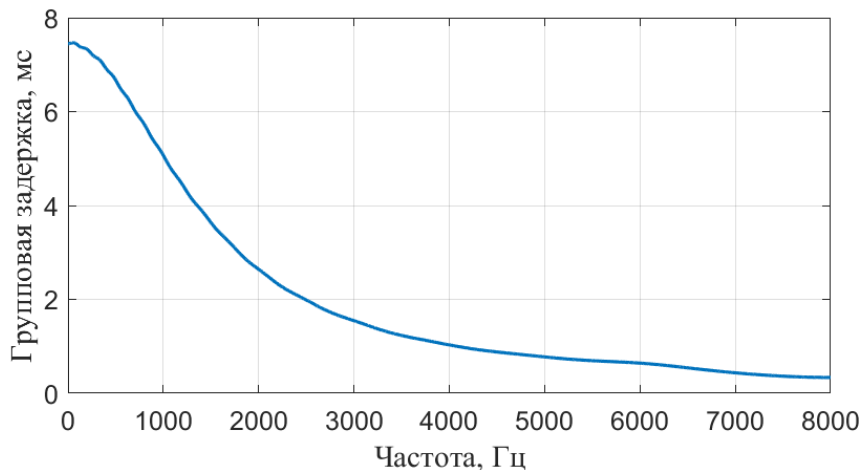


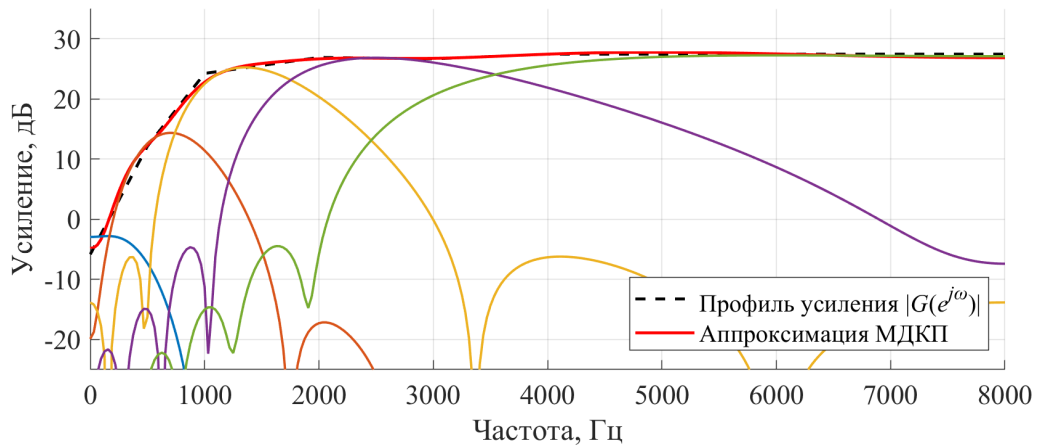
Рисунок 9. – Групповая задержка 22-канальной системы обработки на основе МДКП с деформированной частотной осью

Для аппроксимации профиля усиления $|G(e^{j\omega})|$ коэффициенты g_k в системе обработки (см. рисунок 8) находятся путем минимизации функции потерь:

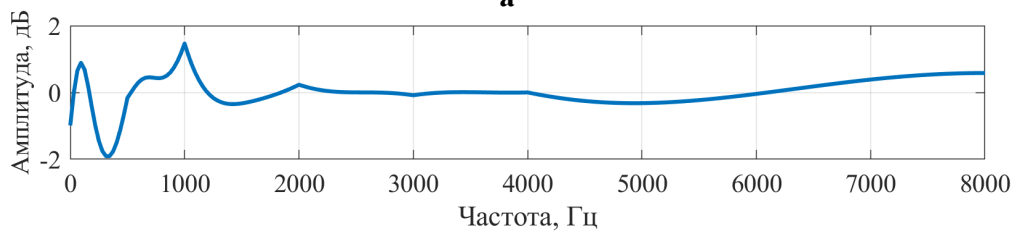
$$\epsilon = \frac{1}{N_{pt}} \sum_{i=0}^{N_{pt}-1} \left| 20 \log_{10}(|G(e^{j\omega_i})|) - 20 \log_{10} \left(\left| \sum_{k=0}^{M-1} g_k H_k(e^{j\omega_i}) \right| \right) \right|, \quad (20)$$

где $\omega_i \in [0, \pi]$ – частотная сетка; N_{pt} – число точек частотной сетки.

На рисунке 10 показан результат подгонки коэффициентов g_k в 5-канальной системе коррекции слуха на основе МДКП с деформированной частотной осью (на рисунке 10,б показана ошибка аппроксимации профиля усиления).



а



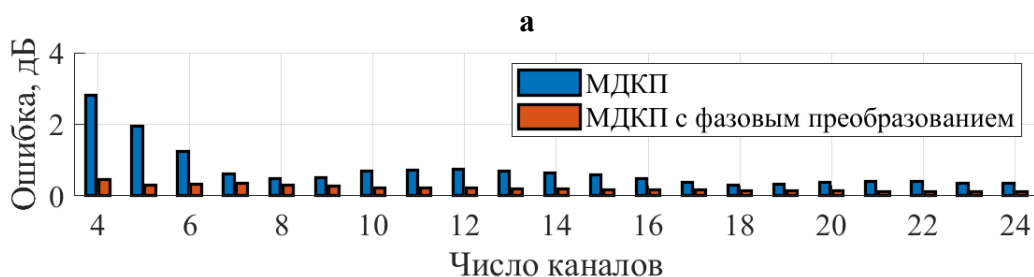
б

а – АЧХ; б – ошибка аппроксимации

Рисунок 10. – Аппроксимация профиля усиления для 5-канальной системы обработки сигнала на основе МДКП с деформированной частотной осью

В главе четыре также приводятся результаты экспериментов по оценке точности аппроксимации профиля усиления при помощи системы на основе МДКП с деформированной частотной осью (см. рисунок 8). В качестве альтернативы рассматривалась система на основе обычного МДКП. Точность аппроксимации (см. выражение (20)) измерялась в зависимости от числа каналов в системе M . Очевидно, что увеличение числа каналов означает и увеличение вычислительных затрат на реализацию системы. Именно поэтому желательно иметь систему с небольшим числом каналов.

В экспериментах использовались различные типы аудиограмм, отражающие различные виды потери слуха. На рисунке 11 показаны ошибки аппроксимации профилей усиления для двух типов аудиограмм.



а – аудиограмма 1 (плоский тип); **б** – аудиограмма 2 (потеря слуха на высоких частотах с умеренным наклоном)

Рисунок 11. – Ошибка аппроксимации профиля усиления, полученного на основе аудиограммы

Из графиков, приведенных на рисунке 11, можно сделать вывод, что использование МДКП с деформированной частотной осью позволяет более эффективно аппроксимировать частотный профиль вносимого усиления, чем обычное МДКП. Например, для того чтобы аппроксимировать профиль частотного усиления, полученный для аудиограммы 1 со средней ошибкой менее 1 дБ, требуется 9-канальный банк фильтров на основе МДКП (ошибка аппроксимации 0,77 дБ), при этом 4-канальный МДКП банк фильтров на основе фазового преобразования имеет ошибку аппроксимации 0,59 дБ. Таким образом, использование банка фильтров на основе МДКП с фазовым преобразованием позволяет получать более точную аппроксимацию профиля вносимого усиления при меньшем числе каналов по сравнению с использованием обычного МДКП.

Метод коррекции слуха на основе МДКП с деформированной частотной осью реализован в виде приложения для мобильной вычислительной платформы. Среди слабослышащих пользователей приложения проводился опрос (на базе анкеты COSI). Результаты показали, что среди пользователей, не носивших ранее слухового аппарата, применение мобильного приложения имеет положительный эффект. В среднем разница между теми, кто понимал более 75 % информации без коррекции, и теми, кто воспринимал более 75 % информации с коррекцией, составила 15,7 %. Среди респондентов, имевших опыт ношения слухового аппарата, средний процент пользователей, которые понимали более 75 % слуховой информации с коррекцией слуха посредством слухового аппарата равен 42,1 %, а в случае использования разработанного приложения этот процент равен 44,4 %.

Система коррекции слуха на основе МДКП не предполагает наличие во входном сигнале помехи от акустической обратной связи (АОС). Обратная связь действительно не возникает, если микрофон и динамик системы разнесены на достаточно большое расстояние. Однако проблема АОС встает очень остро, если микрофон и динамик системы расположены очень близко, как в случае использования беспроводных наушников. Для преодоления ограничений, связанных с возникновением АОС, в главе четыре предложена схема субполосной обработки сигнала на основе ДПФ-модулированного банка фильтров (рисунок 12) [16, 19, 22, 38, 39].

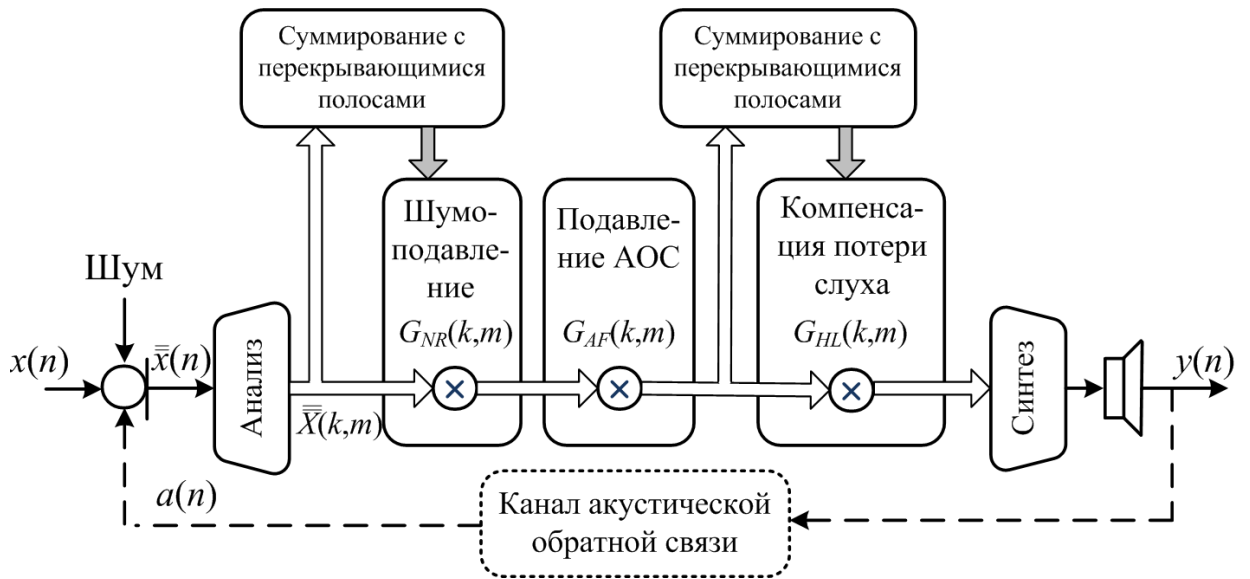


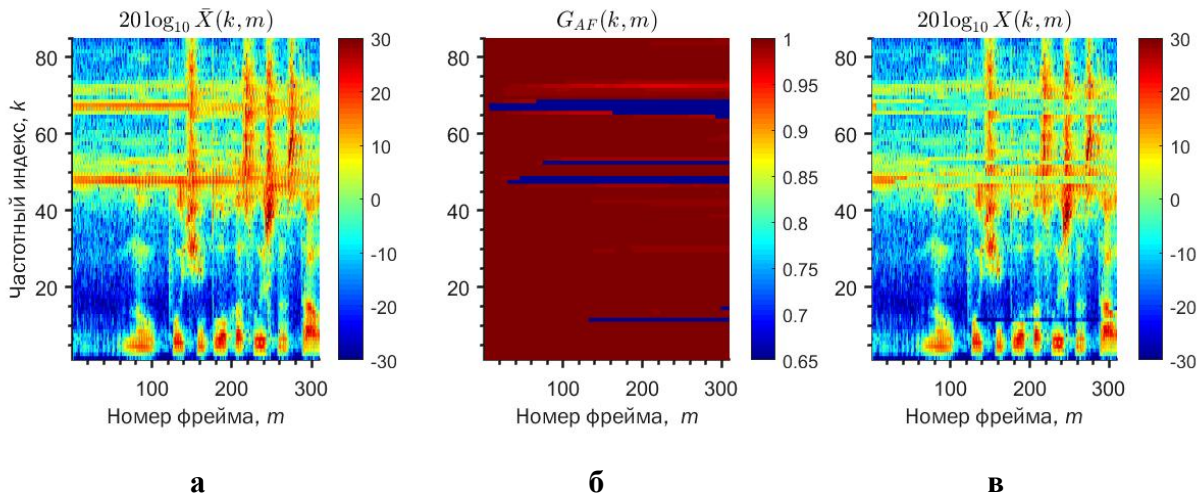
Рисунок 12. – Субполосная схема обработки сигнала

Субполосная система, приведенная на рисунке 12, реализует предложенный в диссертационной работе метод подавления АОС, основанный на принципе спектрального вычитания. При помощи банка фильтров формируется частотно-временное представление сигнала $\bar{X}(k, m)$, где k и m это индексы частоты и времени соответственно. Коэффициенты подавления АОС $\hat{G}_{AF}(k, m)$ вычисляются на основе оценки сигнала АОС:

$$\hat{G}_{AF}(k, m) = \frac{\min_{-d \leq i \leq d} \left\{ \max_{-l+1 \leq j \leq 0} \{ |\bar{X}(k+i, m+j)| \} \right\}}{\min_{-l+1 \leq j \leq 0} \{ |\bar{X}(k, m+j)| \}}, \quad (21)$$

где $|\bar{X}(k, m)|$ – результат очистки входного сигнала от шума:

$$|\bar{X}(k, m)| = |\bar{X}(k, m)| G_{NR}(k, m).$$



а – спектрограмма сигнала, содержащего компоненты АОС; **б** – коэффициенты подавления АОС; **в** – спектрограмма сигнала, очищенного от компонент АОС

Рисунок 13. – Процесс субполосного подавления АОС

Формирование коэффициентов подавления АОС $G_{AF}(k, m)$ выполняется путем рекурсивного усреднения:

$$G_{AF}(k, m) = \begin{cases} \alpha_{AF} G_{AF}(k, m-1) + (1 - \alpha_{AF}) \min(\hat{G}_{AF}(k, m), 1), & \hat{G}_{AF}(k, m) > \frac{1}{\chi_{th}}, \\ \hat{G}_{AF}(k, m), & \hat{G}_{AF}(k, m) \leq \frac{1}{\chi_{th}}, \end{cases} \quad (22)$$

где χ_{th} – пороговое значение для включения режима сильного подавления АОС; α_{AF} – константа, задающая время усреднения.

На рисунке 13 показан пример, иллюстрирующий работу предложенного субполосного метода подавления АОС.

Разработанная субполосная система подавления АОС сравнивалась с широкополосной системой компенсации АОС на основе адаптивной фильтрации. Общая схема проведения эксперимента в среде MatLab представлена на рисунке 14.



Рисунок 14. – Схема тестирования системы подавления АОС

Для оценки качества работы и сравнения систем подавления АОС использовался объективный параметр сегментированного отношения сигнал/шум (СОСШ), который отражает уровень остаточных компонент АОС и прочих помех, возникающих в результате работы системы подавления АОС.

Таблица 1. – Значения СОСШ и PESQ для систем подавления АОС

ΔG , дБ	СОСШ, дБ		PESQ	
	Адаптивная фильтрация	Спектральное вычитание	Адаптивная фильтрация	Спектральное вычитание
1	9,28	20,89	3,67	4,23
3	9,19	19,64	3,65	4,20
5	9,11	18,97	3,63	4,09
7	9,03	18,54	3,59	4,06
9	8,47	17,44	3,39	4,01
11	7,78	17,08	3,09	3,99
13	7,48	16,10	2,96	3,99
15	6,02	13,20	2,48	3,36
17	1,46	14,16	1,94	3,79

В эксперименте для всех частотных полос использовался единый коэффициент компенсации потери слуха, который задавался как $G_{HL} = MSG + \Delta G$, где MSG – такое максимальное усиление, при котором система с отключенным блоком подавления АОС остается стабильной (не наступает режим самовозбуждения); ΔG добавочное усиление, которое применяют при включенной системе подавления АОС. Для перцептуальной оценки качества выходного сигнала использовалась оценка PESQ, которая находится в пределах от $-0,5$ до $4,5$. Наивысшая оценка показывает, что тестовый сигнал не содержит слышимых искажений.

В таблице 1 приведены полученные значения СОСШ при использовании различных добавочных усилений ΔG (от 1 до 17 дБ). Согласно таблице 1 субполосная система на основе спектрального вычитания позволяет получить более высокое отношение сигнал/шум. Средняя разница параметра СОСШ для моделируемых систем равна 9,80 дБ. В таблице 1 также приведены результаты экспериментов по измерению показателя перцептуального качества PESQ для моделируемых систем подавления АОС. Приведенные данные показывают, что в среднем значение PESQ для субполосной системы выше на 0,81 единицы, чем для системы с адаптивной фильтрацией.

В **пятой главе** описывается разработанная система (рисунок 15) автоматизированного повышения качества восприятия звуковых сигналов в шумовой обстановке [10, 13, 22, 55]. Часто прослушиванию звука при помощи мобильного устройства мешает внешний акустический шум. При этом интенсивность шума бывает так высока, что возникает эффект маскирования полезного сигнала шумом. С целью повышения качества восприятия сигнала предлагается выполнять нелинейное усиление частотно-временных компонент воспроизводимого сигнала, зависящее от спектральных характеристик внешнего шума.

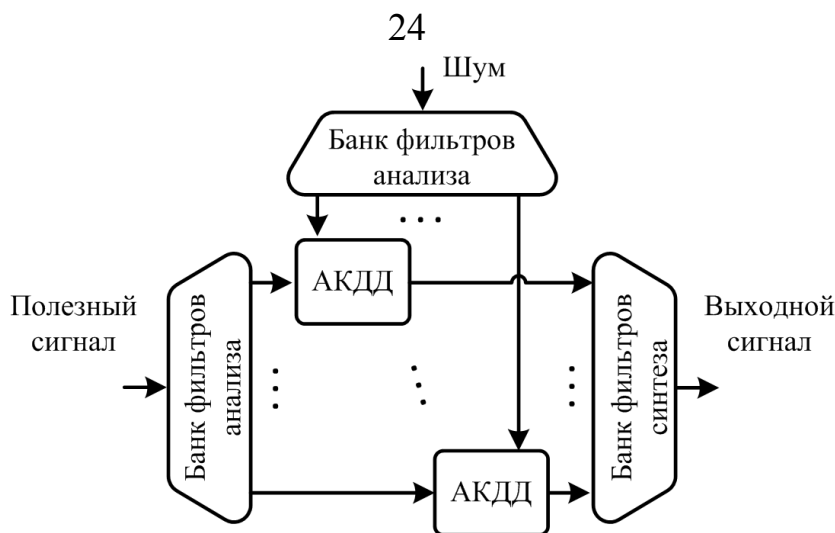


Рисунок 15. – Схема обработки сигнала для повышения качества его воспроизведения в шумовой обстановке

Предложенная схема обработки (см. рисунок 15) выполняет следующие действия: декомпозицию полезного сигнала и шума на субполосы при помощи неравнополосного КМБФ анализа, вычисление кратковременной энергии в каждой субполосе, адаптивную компрессию динамического диапазона (АКДД) каждой субполосы полезного сигнала с учетом кратковременной энергии шума, реконструкцию выходного широкополосного сигнала из обработанных субполосных сигналов при помощи неравнополосного КМБФ синтеза.

Математически АКДД описывается нелинейным оператором:

$$y(n) = \psi \{x(n), w(n)\},$$

где $w(n)$ – шумовой сигнал; $y(n)$ – выходной сигнал компрессора.

При обработке входной и шумовой сигналы разбиваются на фреймы по R отсчетов:

$$x_m(r) = x(mR + r),$$

$$w_m(r) = w(mR + r),$$

где m – номер фрейма; $r = 1, 2, \dots, R$ – индекс отсчета внутри фрейма.

Выходной сигнал АКДД формируется как

$$y_m(n) = x_m(r)g_m(r). \quad (23)$$

Коэффициенты усиления удовлетворяют условию $g^{(m)}(r) \geq 1, \forall r$ и обеспечивают ослабление эффекта маскирования полезного сигнала шумовым.

Предлагаемый в диссертации АКДД имеет амплитудную характеристику (рисунок 16), зависящую от уровня шума W_{nos} и которая определяется двумя параметрами: 1) ΔG определяет динамический диапазон сигналов, к которым приме-

няется усиление и компрессия; 2) ΔR определяет коэффициент компрессии. Особенностью работы адаптивного КДД является то, что в случае отсутствия шума он не вносит искажений в полезный сигнал.

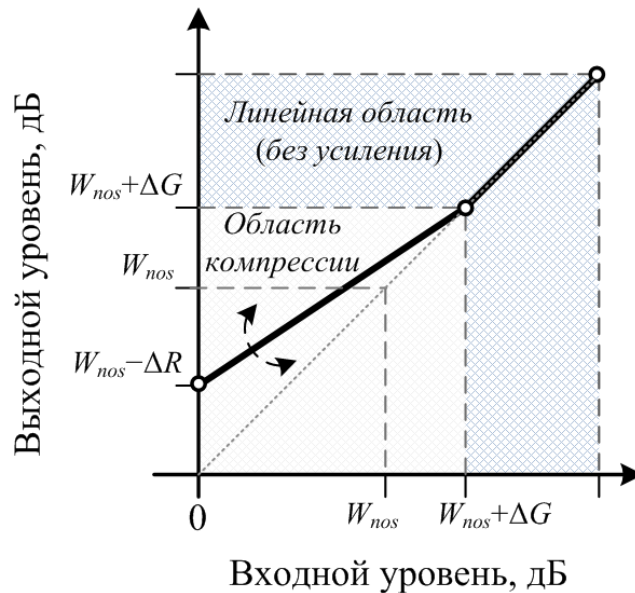


Рисунок 16. – Амплитудная характеристика адаптивного КДД

Объективная оценка метода повышения качества восприятия звукового сигнала выполнена при помощи измерения индекса разборчивости речи (ИРР). Поскольку в рассматриваемой задаче отдельно доступны и полезный сигнал и шумовой, то ИРР можно вычислить достаточно легко. Величина ИРР принимает значение от нуля до единицы. Считается, что ИРР выше 0,75 является хорошим показателем, в то время как значение меньше 0,45 является неудовлетворительным. Речевой материал для выполнения экспериментов взят из свободно распространяемой речевой базы CMU ARCTIC. В качестве шумового сигнала использована запись акустической обстановки в метрополитене, полученная при помощи смартфона. Работа предложенного метода проверялась для различных ОСШ, во всех случаях измерялась разборчивость речевого сигнала на фоне шума до и после обработки. Полученные результаты экспериментов показаны на рисунке 17.

В агрессивной шумовой обстановке (ОСШ < -20 дБ) метод 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.

В **шестой главе** предложен подход к построению системы анализа и классификации голосового сигнала на основе пертурбационных параметров и кепстрального представления. Рассмотрены два варианта кепстрального представления голосового сигнала: при помощи мел-частотных кепстральных коэффициентов (МЧКК) и при помощи барк-частотных кепстральных коэффициентов (БЧКК).

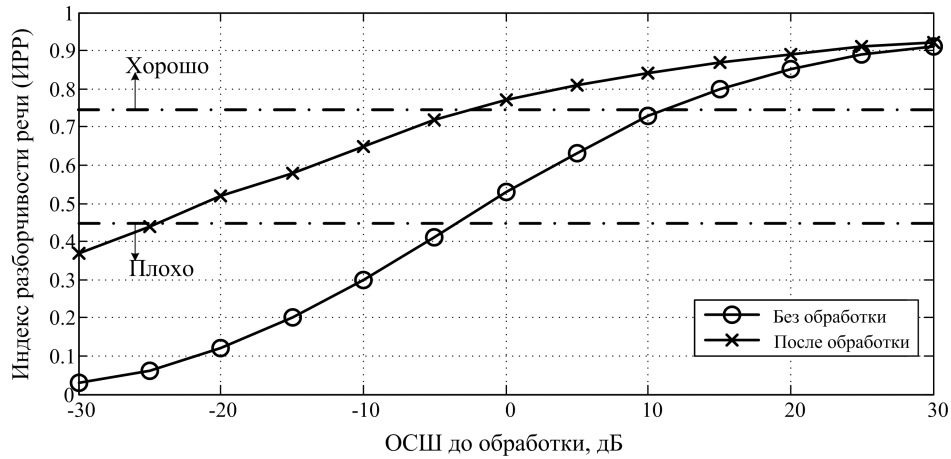


Рисунок 17. – Сравнение среднего значения ИРР для речевых сигналов до и после обработки предложенным методом

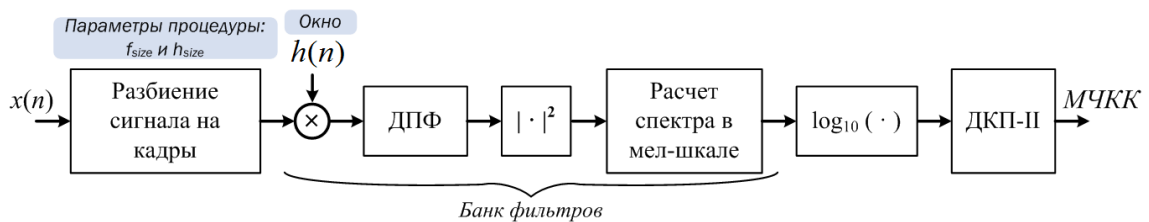


Рисунок 18. – Схема вычисления мел-частотных кепстральных коэффициентов

В работе использовался общепринятый подход к вычислению МЧКК на основе частотно-временного анализа методом дискретного преобразования Фурье (ДПФ) с объединением энергии в субполосах (рисунок 18).

Данный метод аппроксимирует частотное разрешение слуха человека, но имеет фиксированное временное разрешение. В качестве альтернативы предложен вариант кепстрального представления на основе БЧКК [29, 30] (рисунок 19).

При расчете БЧКК использовался неравнополосный ДПФ-модулированный банк фильтров, аппроксимирующий частотную и временную разрешающую способность слуха (рисунок 20).

Целью исследования ставилось сравнение эффективности применения признаков на основе МЧКК и БЧКК для построения систем анализа и классификации голосового сигнала. Результаты эксперимента показали, что в случае использования акустических признаков на основе МЧКК можно получить систему классификации голоса со средней полнотой классификации 80,6 %, а в случае использо-

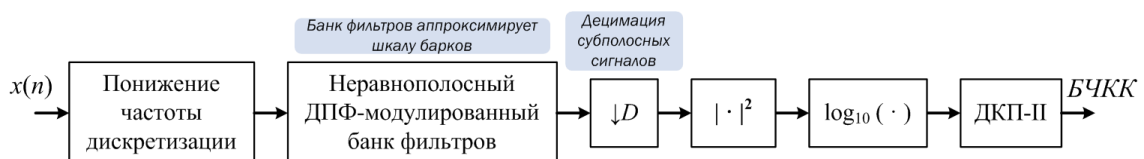
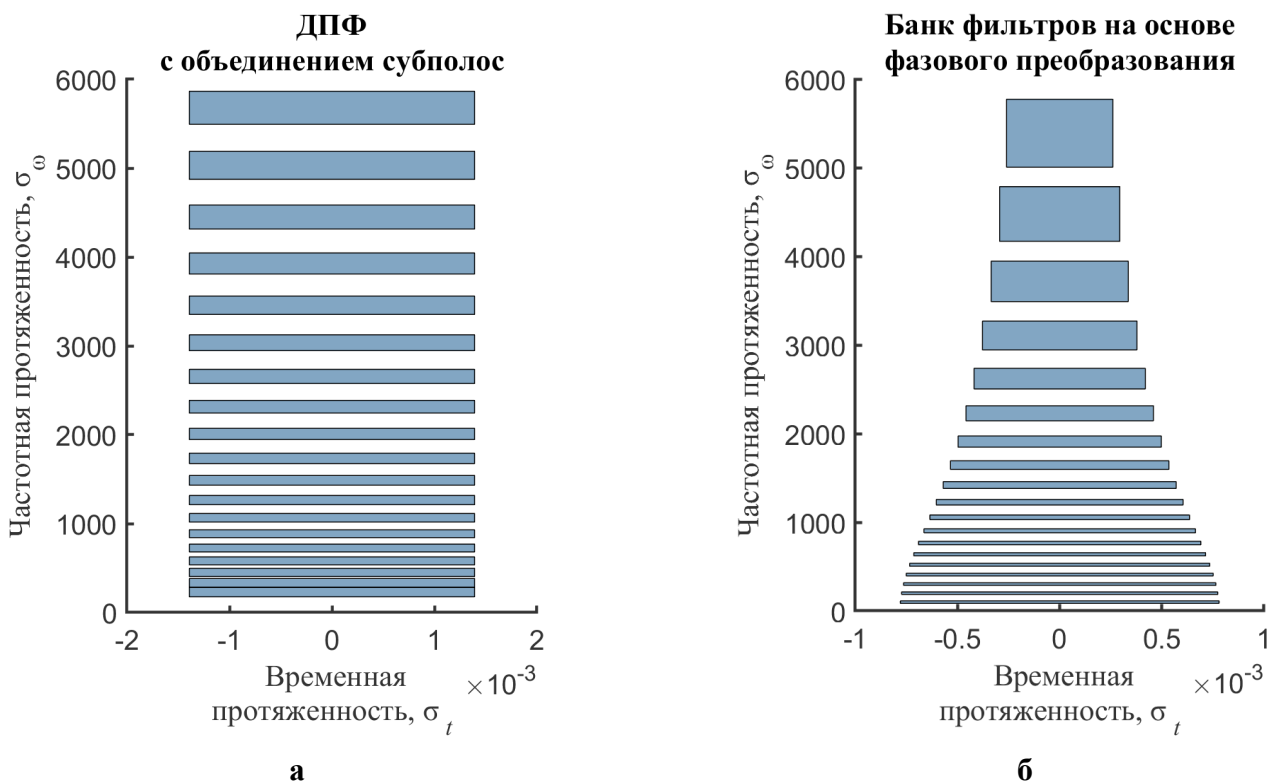


Рисунок 19. – Схема вычисления барк-частотных кепстральных коэффициентов



а – МЧКК; б – БЧКК

Рисунок 20. – Частотно-временные покрытия банков фильтров, применяемых при расчете кепстральных коэффициентов

вании признаков на основе БЧКК этот показатель равен 83,7 %. При дополнении набора МЧКК-признаков пертурбационными параметрами голоса средняя полнота классификации повысилась до 94,1 %, при аналогичном дополнении набора БЧКК-признаков средняя полнота классификации увеличилась до 96,7 %. Для примера на рисунке 21 представлены результаты оценки качества классификации при использовании МЧКК- и БЧКК-признаков.

Помимо кепстральных признаков для построения системы анализа и классификации голосового сигнала используются пертурбационные параметры, получающиеся в результате анализа контура частоты основного тона (ЧОТ): джиттер, шиммер, PPE, DFP и пр. Для анализа голосов пациентов с БАС предложен новый пертурбационный параметр – индекс патологичности вибрато (PVI) [25, 31, 41]. Алгоритм расчета PVI состоит из пяти шагов: 1) расчет контура частоты основного тона (ЧОТ) $f_0(m)$; 2) нормирование контура ЧОТ $f'_0(m) = f_0(m)/E\{f_0(m)\}$; 3) фильтрация $f'_0(m)$ с полосовым фильтром Баттерворта с полосой пропускания 9–14 Гц; 4) расчет амплитудного спектра $A_{f'_0}(f)$ сигнала $f'_0(m)$ методом Уэлча с временным окном в 1 с и перекрытием 95 %; 5) вычисление индекса патологичности вибрато:

$$PVI = \sum_{f \in [9, 14] \text{ Гц}} A_{f'_0}(f). \quad (24)$$

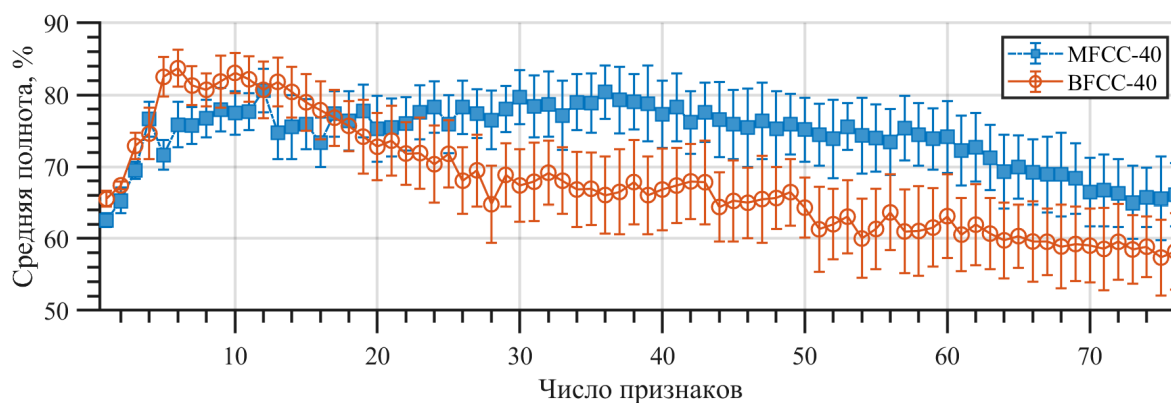


Рисунок 21. – Качество классификации при использовании МЧКК-признаков (MFCC-40) и БЧКК-признаков (BFCC-40)

Для проведения экспериментов использовалась база, содержащая образцы голосовых сигналов (протяжный звук /a/) здоровых людей и пациентов с БАС. База голосов записана в РНПЦ Неврологии и нейрохирургии (г. Минск, РБ), подробная информация о ней, а также ссылка на открытый репозиторий приведены в работе [41].

Результаты экспериментов показали, что применение параметра PVI позволило повысить значение средней полноты классификатора с 87,8 до 90,3 % (т. е. на 2,5 %), при этом общее число пертурбационных признаков было сокращено с десяти до четырех, что является положительным результатом как с точки зрения экономии вычислительных ресурсов, так и с точки зрения упрощения интерпретации процесса классификации.

Седьмая глава посвящена разработке методов анализа модуляционного спектра речевого сигнала [5, 29]. Модуляционный спектр вычисляется при помощи двух последовательных частотно-временных преобразований: первое переводит сигнал в частотную область, второе – в модуляционную. Особенностью предложенного метода является использование неравномерного частотно-временного плана при переводе сигнала в частотную область. Предложен способ оценки модуляционного спектра, основанный на принципах антропоморфической обработки сигналов, имитирующий неравномерное частотно-временное разрешение слуховой системы человека (рисунок 22). Использован неравнополосный ДПФ-модулированный банк фильтров, аппроксимирующий шкалу критических частот и повышающий временное разрешение в широких полосах. Выполнено сравнение предложенного способа с более традиционным способом вычисления модуляционного спектра в критических полосах с фиксированным временным разрешением.

В исследовании строился классификатор для выявления патологии гортани на основе анализа голосового сигнала (протяжного произнесения звука /a/). Использовалась база голосов из работы [27], содержащая 30 записей голосов с

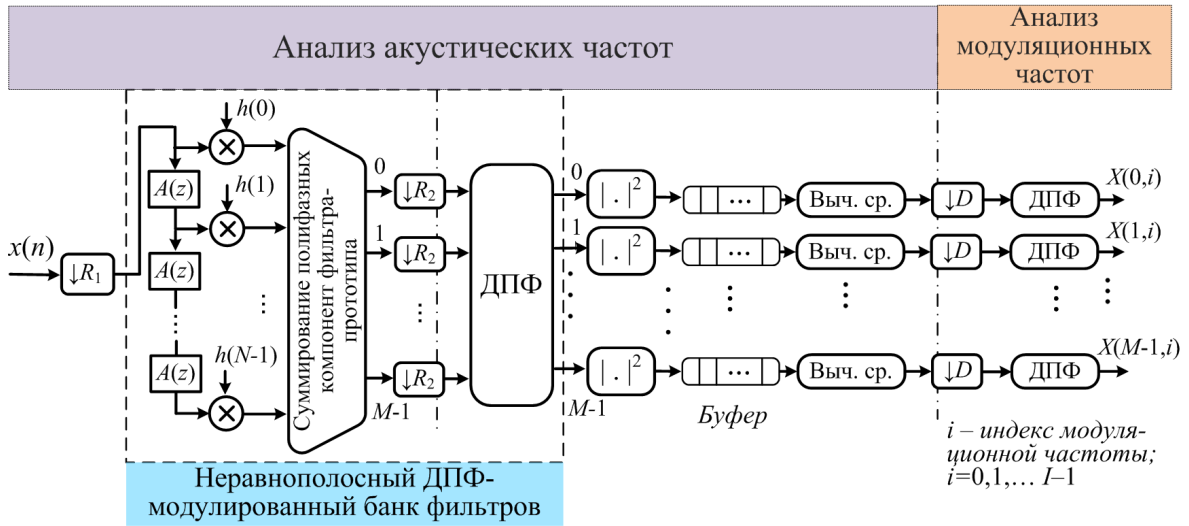


Рисунок 22. – Схема анализа модуляционного спектра сигнала: анализ акустических частот в шкале барков выполняется при помощи неравнополосного банка фильтров

различными видами патологии гортани и 30 записей голосов здоровых людей. В результате перевода звуковой базы в модуляционные параметры получена выборка из 280 обучающих векторов (185 – группа здоровых, 95 – группа с патологиями гортани). На рисунке 23 приведен график, показывающий зависимость правильности классификации от числа признаков и метода анализа модуляционного спектра. В таблице 2 приведены численные значения правильности классификации в зависимости от числа используемых признаков p (в скобках указано число признаков, при котором достигается указанная правильность классификации).

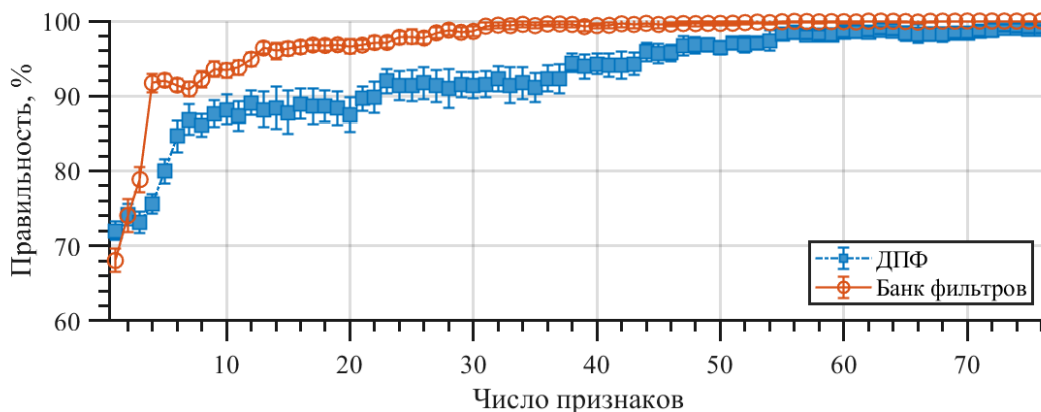


Рисунок 23. – Зависимость правильности классификации от числа признаков

Анализируя график на рисунке 23, можно заметить, что признаки, полученные в схеме на основе неравнополосного банка фильтров, во всех случаях обеспечивают более высокое качество классификации.

В экспериментах по использованию параметров модуляционного спектра для построения классификаторов патологических изменений в голосе использовались базы голосов пациентов: с БАС и с заболеваниями гортани. Отбор и ран-

Таблица 2. – Правильность классификации

Извлечение признаков	Правильность классификации, %						
	$p \leq 10$	$p \leq 20$	$p \leq 30$	$p \leq 40$	$p \leq 50$	$p \leq 60$	$p \leq 70$
Схема на основе ДПФ	88,2 (10)	89,1 (12)	92,0 (23)	94,4 (38)	96,8 (49)	98,5 (60)	98,7 (63)
Схема на основе банка фильтров	93,6 (9)	96,8 (19)	98,8 (28)	99,6 (37)	99,7 (44)	99,9 (56)	100,0 (70)

жирование информационных признаков выполнен при помощи метода LASSO. Достигаемая в результате правильность классификации на основе линейного дискриминантного анализа во всех случаях превышает 97 %. Полученные результаты свидетельствуют о том, что использование неравномерного частотно-временного плана является предпочтительным в случае, когда анализируемый сигнал является протяжным гласным звуком, поскольку обеспечивает более высокую точность обнаружения патологии при меньшем числе модуляционных параметров.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1. Предложен новый подход к построению моделей слухового восприятия на основе неравнополосных банков фильтров с фазовым преобразованием [1, 9]. Показано, что частотно-временное покрытие неравнополосного банка фильтров на основе фазового преобразования имеет сходство с моделью слуховых гамматон-фильтров [42, 49, 52, 54], что позволяет использовать его в качестве метода частотной декомпозиции сигнала для моделирования слухового восприятия [26, 34]. Кроме того, неравнополосные банки фильтров с фазовым преобразованием имеют эффективные алгоритмы реализации, что является преимуществом по отношению к известной модели гамматон-фильтров [2, 6, 10, 29, 30, 32, 36, 42].

2. Разработан новый метод коррекции слухового восприятия, ориентированный на мобильные вычислительные платформы и основанный на частотно-временной декомпозиции с использованием модифицированного дискретного косинусного преобразования с деформированной частотной осью и суммировании субканальных сигналов. Показано, что метод обладает малой алгоритмической задержкой и обеспечивает коррекцию слуха сравнимую с современными цифровыми слуховыми аппаратами [1, 4, 8, 17, 23, 33, 38, 39, 43, 45, 46, 47, 50, 56].

3. Предложен метод подавления акустической обратной связи для системы коррекции слухового восприятия на основе мобильной платформы, основанный на модифицированном правиле спектрального вычитания. Особенностью метода является его высокая устойчивость к резким изменениям характеристик пути распространения обратной связи [16, 19, 38, 39].

4. Разработан оригинальный метод повышения качества восприятия воспро-

изводимого звукового сигнала в шумовой обстановке, основанный на неравнополосном косинусно-модулированном банке фильтров с фазовым преобразованием и правиле нелинейного усиления частотно-временных компонент воспроизводимого сигнала, зависящего от спектральных характеристик окружающего шума. Показано, что метод универсален по отношению к различным типам звуковых сигналов [9, 10, 12, 13, 22, 55].

5. Предложен оригинальный метод анализа модуляционного спектра звукового сигнала, основанный на частотно-временной декомпозиции сигнала неравнополосным банком фильтров с фазовым преобразованием. Показано, что использование его в системах детектирования патологии голосового аппарата позволяет обеспечить в ряде случаев бóльшую правильность классификации, чем применение известного метода анализа на основе ДПФ с объединением энергий в субполосах [5, 29, 54].

6. Разработан алгоритм извлечения из голосового сигнала параметра патологичности вибрато, учитывающий особенности изменения голоса у пациентов с боковым амиотрофическим склерозом (БАС) [28]. Алгоритм основан на количественной оценке частотных компонент вибрато голосового сигнала, характерных для пациентов с БАС [41]. Показано, что его использование в системах детектирования БАС-патологий значительно повышает чувствительность и специфичность системы [25, 31, 35, 53].

7. Предложен метод вычисления барк-частотных кепстральных коэффициентов (БЧКК), основанный на использовании неравнополосного ДПФ-модулированного банка фильтров и разработана оригинальная система классификации голосов пациентов с БАС, основанная на совместном применении надсегментных БЧКК-признаков и пертурбационных параметров голоса (джиттер, шиммер, энтропия периодов основного тона и пр.) [3, 7, 11, 14, 15, 18, 20, 21, 24, 27, 37, 40, 44, 48], что позволяет повысить среднюю полноту классификации до уровня 96,7 % [31, 35, 41, 51].

Рекомендации по практическому использованию результатов

Разработаны и внедрены на предприятиях следующие способы обработки звуковой информации и программные средства:

1. Метод повышения разборчивости звуковых сигналов с учетом персональных слуховых особенностей пользователя и окружающей шумовой обстановки. Метод использует аудиометрические данные пользователя для коррекции персональных особенностей слуха и микрофонный вход для регистрации окружающего шума. Метод основан на параметрической компрессии динамического диапазона полезного сигнала в частотных субполосах. Внедрены и используются в частном предприятии «НТЛаб-системы», г. Минск, Республика Беларусь.

2. Метод обработки сигнала с использованием банка фильтров на основе модифицированного дискретного косинусного преобразования с деформированной частотной осью для системы коррекции слуха на базе мобильного телефона. Метод реализован в виде программного модуля. Основным преимуществом метода является возможность адаптивного выбора числа каналов банка фильтров в зависимости от типа аудиограммы пользователя. В системе выбирается минимальное число каналов, обеспечивающее аппроксимацию профиля усиления сигнала с заданной точностью. Программный модуль внедрен и используется в программном продукте для коррекции слуха «Petralex Hearing Aid App». – IT FORYOU, г. Род-Таун, Британские Виргинские Острова.

3. Метод субполосного подавления сигнала акустической обратной связи, основанный на принципе спектрального вычитания и устойчивый к резким изменениям характеристик пути распространения сигнала обратной связи. Метод реализован в виде программного модуля. Основным преимуществом модуля является его эффективность подавления акустической обратной связи, обеспечивающая высокое перцептуальное качество выходного сигнала. Модуль внедрен в программный продукт для коррекции слуха «Petralex Hearing Aid App» – IT FORYOU, г. Род-Таун, Британские Виргинские Острова.

4. Технология обработки речевых сигналов входных и выходных каналов связи операторов контактных центров в реальном масштабе времени. Технология позволяет улучшать восприятие голоса оператора удаленным абонентом и повышать разборчивость речи удаленного абонента для оператора. Для достижения требуемого эффекта используется моделирование процесса слухового восприятия. Технология внедрена в программный продукт «Petralex Speech Communication» – ООО «АйТиФо Ю», г. Москва, Российская Федерация.

5. Разработанные методы анализа голоса использовались для построения системы детектирования изменений в голосе пациентов с БАС. Система детектирования была реализована в виде приложения для мобильного телефона и тестировалась в РНПЦ «Неврологии и нейрохирургии» (г. Минск, Республика Беларусь). Предварительные результаты тестирования показали, что разработанная система обладает высокой специфичностью.

Благодарности

Автор выражает признательность первому научному консультанту профессору д.т.н. Петровскому А.А. за создание и плодотворное развитие научной школы, а также за всестороннюю личную поддержку.

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ УЧЕНОЙ СТЕПЕНИ**Монография**

1. Вашкевич, М. И. Косинусно-модулированные банки фильтров с фазовым преобразованием: реализация и применение в слуховых аппаратах / М. И. Вашкевич, И. С. Азаров, А. А. Петровский – М. : Горячая линия – Телеком, 2014. – 210 с.

Статьи в научных журналах

2. Процессор ДПФ-модулированного банка фильтров с объединением субканалов для задач мониторинга широкого частотного диапазона / М. И. Вашкевич, А. А. Петровский, Д. И. Каплун, Д. М. Клионский // Известия высших уч. заведений России. Радиоэлектроника. – 2012. – № 6. – С. 53–64.

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

4. Parfieniuk, M. Short-critical-path and structurally orthogonal scaled CORDIC-based approximations of the eight-point discrete cosine transform / M. Parfieniuk, M. Vashkevich, A. Petrovsky // IET Circuits, Devices & Systems. – 2013. – Vol. 7, № 3. – P. 150–158.

5. Алгоритм очистки речевого сигнала от сложных помех путем фильтрации в модуляционной области / И. С. Азаров, М. И. Вашкевич, Д. С. Лихачев, А. А. Петровский // Цифровая обработка сигналов. – 2013. – № 1. – С. 3–10.

6. Изменение частоты основного тона речевого сигнала на основе гармонической модели с нестационарными параметрами / И. С. Азаров, М. И. Вашкевич, Д. С. Лихачев, А. А. Петровский // Труды СПИИРАН. – 2014. – № 1. – С. 5–26.

7. Азаров, И. С. Конверсия голоса в реальном масштабе времени при помощи искусственной нейронной сети с кусочно-линейной функцией активации / И. С. Азаров, М. И. Вашкевич, А. А. Петровский // Нейрокомпьютеры. – 2014. – № 5. – С. 10–18.

8. Система коррекции слуха на мобильной вычислительной платформе / И. С. Азаров, М. И. Вашкевич, С. В. Козлова, А. А. Петровский // Информатика. – 2014. – Т. 42, № 2. – С. 6–25.

9. Анализаторы речевых и звуковых сигналов: методы, алгоритмы и реализации / А. А. Петровский, И. С. Азаров, Ал. А. Петровский, М. И. Вашкевич, Н.А. Петровский // Наука – инновационному развитию общества. Беларуская навука. – 2014. – С. 93–129.

10. Повышение качества слухового восприятия в шумовой обстановке на основе частотной декомпозиции и адаптивной компрессии динамического диапазона сигнала / И. С. Азаров, М. И. Вашкевич, Д. С. Лихачев, А. А. Петровский // Цифровая обработка сигналов. – 2015. – № 1. – С. 3–10.

11. Азаров, И. С. Оценка мгновенных параметров сигнала при помощи модифицированного метода Прони / И. С. Азаров, М. И. Вашкевич, А. А. Петровский // Автоматика и вычислительная техника. – 2015. – № 2. – С. 65–80.

12. Azarov, E. Speech enhancement in quasi-periodic noises using improved spectral subtraction based on adaptive sampling / E. Azarov, M. Vashkevich, A. Petrovsky // Speech and computer, Lecture notes in computer science. – 2015. – Vol. 9319. – P. 405–412.

13. General-purpose listening enhancement based on subband nonlinear amplification with psychoacoustic criterion / E. Azarov, M. Vashkevich, V. Herasimovich, A. Petrovsky // AES Convention paper 9265. – 2015. – P. 10.

14. Azarov, E. Speech analysis based on sinusoidal model with time-varying parameters / E. Azarov, M. Vashkevich, A. Petrovsky // AES Convention paper 9267. – 2015. – P. 8.

15. A Low-Delay Algorithm for Instantaneous Pitch Estimation / E. Azarov, M. Vashkevich, D. Likhachov, A. Petrovsky // AES Convention paper 9268. – 2015. – P. 9.

16. Вашкевич, М. И. Алгоритм подавления шума и акустической обратной связи на основе спектрального вычитания в слуховом протезе на базе смартфона / М. И. Вашкевич, И. С. Азаров, А. А. Петровский, // Доклады БГУИР. – 2016. – № 7. – С. 313–317.

17. Козлова, С. В. Применение мобильного приложения Petralex для коррекции слуховых нарушений / С.В. Козлова, М. И. Вашкевич // Оториноларингология. Восточная Европа. – 2018. – Т. 8, № 1. – С. 81–89.

18. Вашкевич, М. И. Оценка мгновенной частоты основного тона речевого сигнала на основе многоскоростной обработки / М. И. Вашкевич, И. С. Азаров, А. А. Петровский // Речевые технологии. – 2018. – № 1. – С. 12–24.

19. Вашкевич, М. И. Speech enhancement in a smartphone-based hearing aid / М. И. Вашкевич, И. С. Азаров, А. А. Петровский // Речевые технологии. – 2018. – № 1. – С. 63–81.

20. Гвоздович, А. Д. Детектирование бульбарных нарушений при боковом амиотрофическом склерозе на основе анализа речевого сигнала / А. Д. Гвоздович, Ю. Н. Рушкевич, М. И. Вашкевич // Доклады БГУИР. – 2018. – № 6. – С. 52–58.

21. Гвоздович, А. Д. Акустический анализ голоса для выявления речевых нарушений при боковом амиотрофическом склерозе / М. И. Вашкевич, А. Д. Гвоздович, Ю. Н. Рушкевич, А. А. Петровский // Доклады БГУИР. – 2018. – № 7. – С. 64–68.
22. Вашкевич, М. И. Цифровые банки фильтров для современных задач обработки звуковых сигналов / М. И. Вашкевич, И. С. Азаров, А. А. Петровский // Доклады БГУИР. – 2019. – № 2. – С. 100–111.
23. Аппаратный аудиологический скрининг детей младшего школьного возраста в Гродненской области с использованием программного обеспечения PetraleX / О. Г. Хоров, Д. Н. Марцуль, Е. И. Никита, С. Н. Ракова, Е. И. Головач, А. А. Петровский, М. И. Вашкевич // Оториноларингология. Восточная Европа. – 2019. – Т. 9, № 2. – С. 166–178.
24. Vashkevich, M. Detection of bulbar dysfunction in ALS patients based on running speech test / M. Vashkevich, A. Gvozдович, Y. Rushkevich // Pattern Recognition and Information Processing, Communications in Computer and Information Science. – Vol. 1055. – 2019. – P. 192–204.
25. Вашкевич, М. И. Детектирование речевых нарушений у пациентов с БАС на основе анализа голосового сигнала / М. И. Вашкевич, Ю. Н. Рушкевич // Цифровая обработка сигналов. – 2019. – № 4. – С. 33–41.
26. Порхун, М. И. Метод коррекции слуха на основе психоакустически обусловленного переноса частот в речевом сигнале / М. И. Порхун, М. И. Вашкевич // Доклады БГУИР. – 2020. – № 1. – С. 43–51.
27. Анализ акустических параметров голоса для выявления заболеваний гортани / М. И. Вашкевич, А. А. Бурак, Н. С. Конойко, В. С. Долдова // Информатика. – 2020. – Т. 17, № 1. – С. 78–86.
28. Вашкевич, М. И. Обзор систем автоматического детектирования речевых нарушений у пациентов с боковым амиотрофическим склерозом (БАС) / М. И. Вашкевич, Ю. Н. Рушкевич // Речевые технологии. – 2020. – № 1–2. – С. 73–85.
29. Вашкевич, М. И. Определение патологии голосового аппарата на основе анализа модуляционного спектра речи в критических полосах / М. И. Вашкевич, И. С. Азаров // Труды СПИИРАН. – 2020. – № 2. – С. 249–276.
30. Вашкевич, М. И. Сравнение частотно-временных преобразований: Фурье-анализ, вейвлеты и банки фильтров на основе фазового преобразования / М. И. Вашкевич, И. С. Азаров // Цифровая обработка сигналов. – 2020. – № 2. – С. 13–26.
31. Vashkevich, M. Classification of ALS patients based on acoustic analysis of sustained vowel phonations / M. Vashkevich, Y. Rushkevich // Biomedical Signal Processing and Control. – 2021. – Vol 65. – P.1–14.

32. Вашкевич, М. И. Анализ ошибок округления при реализации цепочки фазовых звеньев в арифметике с фиксированной запятой // Цифровая обработка сигналов. – 2021. – № 2. – С. 12–18.

33. Санько, Н. С. Исследование применимости ДПФ-модулированного банка фильтров в системах со значительным усилением спектральных составляющих / Н. С. Санько, М. И. Вашкевич // Доклады БГУИР. – 2021. – № 6. – С. 14–22.

34. Порхун, М. И. Метод моделирования эффекта ухудшения частотного разрешения слуха у больных нейросенсорной тугоухостью / М. И. Порхун, М. И. Вашкевич // Информатика. – 2021. – Т. 18, № 3. – С. 68–82.

35. Likhachov, D. A Mobile application for detection of amyotrophic lateral sclerosis via voice analysis / D. Likhachov, M. Vashkevich, E. Azarov, K. Malhina, Y. Rushkevich // Speech and Computer. SPECOM 2021. Lecture Notes in Computer Science – Vol. 12997. – 2021. – P. 372–383.

**Статьи в сборниках материалов научных конференций,
включенных в системы международного цитирования
(Scopus, Web of Science и IEEE Xplore digital library)**

36. Vashkevich, M. Multidimensional DFT-modulated filter bank: design and implementation / M. Vashkevich, A. Petrovsky // Audio, language and image processing (ICALIP'2014): proc. of int. conf., Shanghai, China, July 7–9, 2014. – Shanghai, 2014. – P. 686–690.

37. Azarov, E. Instantaneous pitch estimation algorithm based on multirate sampling / E. Azarov, M. Vashkevich, A. Petrovsky // Acoustic, speech and signal processing (ICASSP'2016): proc. of IEEE int. conf., Shanghai, China, 20–25 May, 2016. – Shanghai, 2016. – P. 4970–4974.

38. Real-time implementation of hearing aid with combined noise and acoustic feedback reduction based on smartphone / M. Vashkevich, E. Azarov, N. Petrovsky, D. Likhachov, A. Petrovsky // Acoustic, speech and signal processing (ICASSP'2017): proc. of IEEE int. conf., Louisiana, USA, March 5–9, 2017. – Louisiana, 2017. – P. 6570–6571.

39. Petralex: A smartphone-based real-time digital hearing aid with combined noise reduction and acoustic feedback suppression / M. Vashkevich, E. Azarov, N. Petrovsky, A. Petrovsky // Signal processing algorithms, architectures, arrangements, and applications (SPA'2017): proc. of the IEEE int. conf., Poland, Poznan, Sept. 20–22, 2017. – Poznan, 2017. – P. 249–254.

40. Features extraction for the automatic detection of ALS disease from acoustic speech signals / M. Vashkevich, E. Azarov, A. Petrovsky, Y. Rushkevich // Signal processing algorithms, architectures, arrangements, and applications (SPA'2018): proc. of the IEEE inter. conf., Poland, Poznan, Sept. 19–21, 2018. – Poznan, 2018. – P. 321–326.

41. Vashkevich, M. Bulbar ALS detection based on analysis of voice perturbation and vibrato / M. Vashkevich, Y. Rushkevich, A. Petrovsky // Signal processing algorithms, architectures, arrangements, and applications (SPA'2019): proc. of the IEEE int. conf., Poland, Poznan, Sept. 18–20, 2019. – Poznan, 2019. – P. 267–272.

Статьи в сборниках материалов научных конференций

42. Вашкевич, М. И. Слуховой аппарат на основе кохлеарной модели с субполосным подавлением эффекта обратной связи / М. И. Вашкевич, И. С. Азаров, А. А. Петровский // Цифровая обработка сигналов и ее применение : сб. тр. 15-й Междунар. конф., Россия, Москва, 27–29 марта 2013 г. – Москва, 2013. – Т.1 – С. 132–135.

43. Вашкевич, М. И. Исследование возможности применения модифицированного дискретного косинусного преобразования с деформацией частотной оси для построения слухового аппарата / М. И. Вашкевич, И. С. Азаров, А. А. Петровский // Цифровая обработка сигналов и ее применение : сб. тр. 15-й Междунар. конф., Россия, Москва, 27–29 марта 2013 г. – Москва, 2013. – Т.1 – С. 109–112.

44. Передискретизация речевого сигнала, согласованная с частотой основного тона / М. И. Вашкевич, И. С. Азаров, Д. С. Лихачев, А. А. Петровский // Междунар. науч.-техн. конф., приуроченная к 50-летию МРТИ-БГУИР. Материалы конференции. Беларусь, Минск, 18–19 марта 2014 г. – Минск, 2014. – Т.1 – С. 308–309.

45. Вашкевич, М. И. Реализация слухового аппарата на мобильной вычислительной платформе / М. И. Вашкевич, И. С. Азаров, А. А. Петровский // Цифровая обработка сигналов и ее применение : сб. тр. 16-й Междунар. конф., Россия, Москва, 26–28 марта 2014 г. – Москва, 2014. – Т.1 – С. 151–155.

46. Азаров, И. С. Исследование возможности использования вычислительной платформы iPhone для коррекции слуха / И. С. Азаров, М. И. Вашкевич, А. А. Петровский // Цифровая обработка сигналов и ее применение : сб. тр. 16-й Междунар. конф., Россия, Москва, 26–28 марта 2014 г. – Москва, 2014. – Т.1 – С. 155–159.

47. Реализация слухового протеза на современных мультимедийных платформах для коррекции тугоухости легкой и средней степени / М. И. Вашкевич, И. С. Азаров, Д. С. Лихачев, А. А. Петровский // Медэлектроника 2014 : сб. тр. Междунар. науч.-техн. конф., Беларусь, Минск, 10–11 декабря 2014 г. – Минск, 2014. – С. 147–150.

48. Выделение признаков неврологических заболеваний из речевого сигнала / М. И. Вашкевич, Ю.Н. Рушкевич, И. С. Азаров, А. А. Петровский // Цифровая обработка сигналов и ее применение : сб. тр. 20-й Междунар. конф., Россия, Москва, 28–30 марта 2018 г. – Москва, 2018. – Т.1 – С. 179–184.

49. Порхун М. И. Моделирование эффекта потери слуха / М. И. Порхун, М. И. Вашкевич // Цифровая обработка сигналов и ее применение : сб. тр. 20-й Междунар. конф., Россия, Москва, 28–30 марта 2018 г. – Москва, 2018. – Т.1 – С. 228–233.

50. Вашкевич, М. И. Платформа скринингового обследования слуха у детей раннего школьного возраста на основе смартфона / М. И. Вашкевич, О. Г. Хоров, А. А. Петровский // Медэлектроника 2018 : сб. тр. Междунар. науч.-техн. конф., Беларусь, Минск, 5–6 декабря 2018 г. – Минск, 2018. – С. 25–28.

51. Detection bulbar dysfunction in ALS patients using acoustic analysis of vowels extracted from continuous speech / A. Gvozdoch, M. Vashkevich, Y. Rushkevich, A. Petrovsky // Pattern recognition and information processing (PRIP'2019): proc. of int. conf., Belarus, Minsk, May 21–23, 2019. – Minsk, 2019. – P. 120–124.

52. Порхун М. И. Моделирование нейросенсорной тугоухости на основе субполосной компрессии динамического диапазона / М. И. Порхун, М. И. Вашкевич // Цифровая обработка сигналов и ее применение : сб. тр. 22-й Междунар. конф., Россия, Москва, 25–27 марта 2020 г. – Москва, 2020. – С. 149–153.

53. Оценка степени нарушения голосовой функции у пациентов с боковым амиотрофическим склерозом с использованием мобильного приложения / Д. С. Лихачев, Е. С. Наркевич, Ю. Н. Рушкевич, М. И. Вашкевич // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня: сб. тр. 6-й Междунар. конф., Беларусь, Минск, 20–21 мая 2020 г. – Минск, 2020. – ч.2. – С. 102–113.

54. Вашкевич, М. И. Биоинспирированный и психоакустически мотивированный анализ звуковых сигналов / М. И. Вашкевич, И. С. Азаров // Информационные технологии и системы 2020 (ИТС 2020) : мат. междунар. науч. конф., Минск, 18 ноября 2020 г. – Минск, 2020. – С. 9–12.

Патенты

55. Способ повышения разборчивости и информативности звуковых сигналов в шумовой обстановке : Пат. 2589298 Российская Федерация : № 2014153295/08 / М. И. Вашкевич, И. С. Азаров, А. Ю. Бредихин, А. А. Петровский; дата публ.: 10.07.16.

56. Способ компенсации потери слуха в телефонной системе и в мобильном телефонном аппарате : Пат. 2568281 Российская Федерация : № 2013125243/08 / А. Ю. Бредихин, М. И. Вашкевич, И. С. Азаров, А. А. Петровский; дата публ.: 10.04.15.



РЕЗЮМЕ

Вашкевич Максим Иосифович

**МЕТОДЫ ЧАСТОТНО-ВРЕМЕННОГО АНАЛИЗА И ОБРАБОТКИ
ЗВУКОВЫХ СИГНАЛОВ НА ОСНОВЕ
БАНКОВ ФИЛЬТРОВ С ФАЗОВЫМ ПРЕОБРАЗОВАНИЕМ**

Ключевые слова: неравнополосные банки фильтров на основе фазового преобразования, частотно-временной анализ, обработка речевых и голосовых сигналов, модель слухового восприятия, повышение разборчивости речи, подавление акустической обратной связи, анализ модуляционного спектра.

Цель работы: разработка методов и алгоритмов частотно-временного анализа звуковых сигналов на основе неравнополосных банков фильтров с фазовым преобразованием и построение на их основе систем обработки звуковой информации.

Методы исследования: в работе использовались методы цифровой обработки сигналов, математического моделирования, машинного обучения, спектрального анализа, методы цифрового моделирования на ЭВМ.

Полученные результаты и их новизна: разработаны методологические основы построения систем обработки звуковой информации на основе частотно-временной декомпозиции сигнала, выполняемой неравнополосным банком фильтров с фазовым преобразованием, позволяющим аппроксимировать частотно-временное разрешение слуха человека; разработаны математические и алгоритмические средства, необходимые для анализа, обработки и реконструкции звукового сигнала и обеспечивающие широкие возможности для создания новых и улучшения существующих систем обработки звуковой информации, включая системы коррекции слуха и повышения качества восприятия, распознавания звуковой информации и детектирования голосовых патологий.

Степень использования: методы обработки звуковой информации, программные модули внедрены и используются на предприятиях: ЧУП «НТЛаб-системы», г. Минск, Республика Беларусь; «IT FOR YOU» г. Род-Таун, Британские Виргинские Острова; «АйТи Фо Ю», г. Москва, Российская Федерация.

Область применения: разработанные методы частотно-временного анализа, синтеза и преобразования звуковых сигналов могут использоваться для решения различных задач обработки звуковой информации, в том числе для распознавания речи, шумоподавления, идентификации диктора, распознавания эмоционального состояния диктора, детектирования заболеваний по голосу.

Вашкевіч Максім Іосіфавіч

МЕТАДЫ ЧАСТОТНА-ЧАСОВАГА АНАЛІЗУ І АПРАЦОЎКІ ГУКАВЫХ
СІГНАЛАЎ НА АСНОВЕ БАНКАЎ ФІЛЬТРАЎ З ФАЗАВЫМ
ПЕРАЎТВАРЭННЕМ

Ключавыя словы: нераўнапалосныя банкі фільтраў на аснове фазовага пераўтварэння, частотна-часавы аналіз, апрацоўка маўленчых і галасавых сігналаў, мадэль слыхавога ўспрымання, павышэнне пераборлівасці гаворкі, прыгнечанне акустычнай зваротнай сувязі, аналіз мадуляцыйнага спектру.

Мэта работы: распрацоўка метадаў і алгарытмаў частотна-часовага аналізу гукавых сігналаў на аснове нераўнапалосных банкаў фільтраў з фазавым пераўтварэннем і пабудова на іх аснове сістэм апрацоўкі гукавой інфармацыі.

Метады даследавання: у працы выкарыстоўваліся метады лічбавай апрацоўкі сігналаў, матэматычнага мадэлявання, машыннага навучання, спектральнага аналізу, метады лічбавага мадэлявання на ЭВМ.

Атрыманыя вынікі і іх навізна: распрацаваны метадалагічныя асновы пабудовы сістэм апрацоўкі гукавой інфармацыі на аснове частотна-часовай дэкампазіцыі сігнала, выкананай нераўнапалосным банкам фільтраў з фазавым пераўтварэннем, якія дазваляюць апраксімаваць частотна-часавы распазнаванне слыху чалавека; распрацаваны матэматычныя і алгарытмічныя сродкі, неабходныя для аналізу, апрацоўкі і рэканструкцыі гукавога сігнала, якія ствараюць шырокія магчымасці для пабудовы новых і паляпшэння існуючых сістэм апрацоўкі гукавой інфармацыі, уключаючы сістэмы карэкцыі слыху і павышэння якасці ўспрымання, распазнавання гукавой інфармацыі і дэтэктавання галасавых паталогій.

Ступень выкарыстання: метады апрацоўкі гукавой інфармацыі, праграмныя модулі ўкаранёны і выкарыстоўваюцца на прадпрыемствах: ПУП «НТЛаб-сістэмы», г. Мінск, Рэспубліка Беларусь; «IT FOR YOU» г. Род-Таун, Брытанскія Віргінскія Астравы; «АйТи Фо Ю», г. Масква, Расійская Федэрацыя.

Вобласць ужывання: распрацаваныя метады частотна-часовага аналізу, сінтэзу і пераўтварэнні гукавых сігналаў могуць выкарыстоўвацца для рашэння розных задач апрацоўкі гукавой інфармацыі, у тым ліку для распазнавання прамовы, шумапрыглушэння, ідэнтыфікацыі дыктара, распазнаванні эмацыйнага стану дыктара, дэтэктаванні захворванняў па голасе.

SUMMARY

Vashkevich Maksim Iosifovich

METHODS OF TIME-FREQUENCY ANALYSIS AND PROCESSING OF AUDIO SIGNALS BASED ON FILTER BANKS WITH ALLPASS TRANSFORM

Key words: nonuniform filter banks based on allpass transform, time-frequency analysis, speech and voice signal processing, auditory perception model, speech intelligibility enhancement, acoustic feedback suppression, modulation spectrum analysis.

The aim of the work: development of methods and algorithms for time-frequency analysis of audio signals based on nonuniform filter banks with allpass transform and construction of audio information processing systems based on them.

Research methods: digital signal processing, mathematical modeling, machine learning, spectral analysis, computer digital modeling.

Results and their novelty: methodological foundations for building sound information processing systems based on time-frequency signal decomposition performed by a nonuniform filter bank with allpass transform which allows approximating the time-frequency resolution of human hearing have been developed; developed mathematical and algorithmic tools necessary for the analysis, processing and reconstruction of the sound signal, and creating wide possibilities for designing new and improving existing systems for processing sound information, including systems for hearing correction and improving the quality of perception, recognizing sound information and detecting voice pathologies.

Usage: sound information processing methods, software modules are implemented and used at enterprises: PUE «NTLab-systems», Minsk, Republic of Belarus, «IT FOR YOU», Road-Town, British Virgin Islands, «ITForYou», Moscow, Russian Federation.

Field of application: developed methods of time-frequency analysis, synthesis and conversion of sound signals can be used to solve various problems of sound information processing, including speech recognition, noise reduction, speaker identification, emotion recognition, voice pathology detection.

Научное издание

Вашкевич Максим Иосифович

**МЕТОДЫ ЧАСТОТНО-ВРЕМЕННОГО АНАЛИЗА И ОБРАБОТКИ
ЗВУКОВЫХ СИГНАЛОВ НА ОСНОВЕ БАНКОВ ФИЛЬТРОВ
С ФАЗОВЫМ ПРЕОБРАЗОВАНИЕМ**

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
доктора технических наук

по специальности 05.13.17 – Теоретические основы информатики

Подписано в печать

Формат 60×84 1/16.

Бумага офсетная.

Гарнитура «Таймс».

Отпечатано на ризографе.

Усл. печ. л.

Уч.-изд. л. 2,7

Тираж 60 экз.

Заказ

Издатель и полиграфическое исполнение: учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники».

Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий № 1/238 от 24.03.2014,

№ 2/113 от 07.04.2014, № 3/615 от 07.04.2014.

ЛП № 02330/264 от 14.04.2014.

Ул. П. Бровки, 6, 220013, г. Минск