Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«Рязанский государственный радиотехнический университет
имени В.Ф. Уткина»**

На правах рукописи



Ву Хоанг Шон

АЛГОРТМЫ МНОГОСКОРОСТНОЙ ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ В РАДИОТЕХНИЧЕСКИХ СИСТЕМАХ ПЕРЕДАЧИ ИНФОРМАЦИИ

Специальность:
2.2.13. – Радиотехника, в том числе системы и устройства телевидения

ДИССЕРТАЦИЯ
на соискание ученой степени кандидата технических наук

Научный руководитель
доктор технических наук, доцент
Дмитриев Владимир Тимурович

Рязань – 2026

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....6

ГЛАВА 1 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ АЛГОРИТМОВ МНОГОСКОРОСТНОЙ ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ НА ОСНОВЕ ПРЕДСТАВЛЕНИЯ ХУРГИНА – ЯКОВЛЕВА.....14

1.1 Вводные замечания 14

1.2 Особенности речевых сигналов и обоснование применимости представления Хургина – Яковлева 16

1.2.1 Статистические и спектральные характеристики речевых сигналов 16

1.2.2 Требования к качеству, скорости передачи и метрики оценки 20

1.3 Теорема отсчётов и обобщённое представление Хургина – Яковлева..... 24

1.3.1 Классическая теорема дискретизации В.А. Котельникова 24

1.3.2 Обобщённое представление Хургина – Яковлева..... 25

1.4 Разработка синтезирующих фильтров для многоканальных систем на основе представления Хургина – Яковлева 28

1.4.1 Двухканальная система на основе представления Хургина – Яковлева. 28

1.4.2 Трёхканальная система на основе представления Хургина – Яковлева 34

1.4.3 Четырёхканальная система на основе представления Хургина – Яковлева 38

1.5 Оценка точности алгоритмов вычисления производных сигнала для систем на основе представления Хургина – Яковлева 42

1.5.1 Требования к алгоритмам вычисления производных..... 42

1.5.2 Методы вычисления производных дискретных сигналов и их применимость при реализации представления Хургина – Яковлева..... 44

1.5.3 Экспериментальные исследования точности алгоритмов вычисления производных 49

1.6 Оценка точности восстановления сигналов на основе представления Хургина – Яковлева 53

1.6.1 Анализ составляющих погрешности восстановления сигнала 53

1.6.2 Методика оценки точности восстановления 55

1.6.3 Экспериментальные исследования точности восстановления 56
1.7 Выводы 60
ГЛАВА 2 РЕАЛИЗАЦИЯ АЛГОРИТМОВ ПЕРВИЧНОГО КОДИРОВАНИЯ РЕЧЕВЫХ СИГНАЛОВ НА ОСНОВЕ ПРЕДСТАВЛЕНИЯ ХУРГИНА – ЯКОВЛЕВА 62
2.1 Вводные замечания 62
2.2 Реализация представления Хургина – Яковлева в виде банка фильтров и оценка вычислительной сложности 64
2.2.1 Эквивалентное представление Хургина – Яковлева в форме многоканального банка фильтров анализа-синтеза 64
2.2.2 Полифазное разложение банка фильтров представления Хургина – Яковлева 68
2.2.3 Оценка вычислительной сложности БПФ 70
2.3 Помехоустойчивость системы обработки речевых сигналов на основе представления Хургина – Яковлева 77
2.3.1 Постановка задачи и модель помех 77
2.3.2 Теоретическая оценка помехоустойчивости при воздействии белого гауссовского шума 79
2.3.3 Экспериментальное исследование помехоустойчивости системы обработки речевых сигналов на основе представления Хургина – Яковлева 84
2.4 Экспериментальное исследование эффективности алгоритмов первичного кодирования речевых сигналов 88
2.4.1 Методика и условия экспериментального исследования 88
2.4.2 Оценка выигрыша качества восстановленной речи в условиях первичного кодирования 90
2.4.3 Сравнительный анализ помехоустойчивости предложенного метода и системы на основе теоремы В.А. Котельникова 93
2.5 Выводы 100

ГЛАВА 3 ИССЛЕДОВАНИЕ ИНТЕГРИРОВАННОЙ СИСТЕМЫ КОДИРОВАНИЯ РЕЧИ НА ОСНОВЕ ПРЕДСТАВЛЕНИЯ ХУРГИНА-ЯКОВЛЕВА И ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ 102

3.1 Вводные замечания 102

3.2 Метод сжатия речевых сигналов на основе покадрового дискретного вейвлет-преобразования 104

3.2.1 Покадровое дискретное вейвлет-преобразование речевых сигналов ... 104

3.2.2 Классические методы пороговой обработки вейвлет-коэффициентов . 107

3.2.3 Алгоритм энергетического порогового сжатия вейвлет-коэффициентов 111

3.2.4 Метрики эффективности сжатия 113

3.3 Оптимизация энергетических порогов вейвлет-преобразования алгоритмом роя частиц 115

3.3.1 Постановка задачи многокритериальной оптимизации 115

3.3.2 Алгоритм роя частиц: структура, параметры и критерии останова 116

3.3.3 Экспериментальная верификация метода: выбор оптимального вейвлет-базиса и результаты оптимизации энергетических порогов 120

3.4 Исследование интегрированной системы кодирования речи на основе представления Хургина – Яковлева и дискретного вейвлет-преобразования ... 129

3.4.1 Структура и параметры интегрированной системы 129

3.4.2 Оптимизация параметров интегрированной системы методом роя частиц PSO 132

3.4.3 Сравнение эффективности интегрированной системы с низкоскоростными кодеками 135

3.5 Предложения по реализации предложенных систем на современной элементной базе 139

3.5.1 Анализ вычислительной сложности интегрированной системы 139

3.5.2 Требования к аппаратным платформам и рекомендации по выбору элементной базы 143

3.6 Выводы 145

ЗАКЛЮЧЕНИЕ **148**

СПИСОК ЛИТЕРАТУРЫ **150**

СЛОВАРЬ СОКРАЩЕНИЙ **164**

ПРИЛОЖЕНИЕ А. ВЫЧИСЛИТЕЛЬНАЯ СЛОЖНОСТЬ СИСТЕМЫ НА ОСНОВЕ ПРЕДСТАВЛЕНИЯ ХУРГИНА – ЯКОВЛЕВА **166**

ПРИЛОЖЕНИЕ Б. КОПИЯ АКТОВ ВНЕДРЕНИЯ **171**

ПРИЛОЖЕНИЕ В. КОПИЯ СВИДЕТЕЛЬСТВ О ГОСУДАРСТВЕННОЙ РЕГИСТРАЦИИ ПРОГРАММЫ ДЛЯ ЭВМ **177**

ПРИЛОЖЕНИЕ Г. КОПИЯ ДИПЛОМОВ **182**

ВВЕДЕНИЕ

Актуальность темы исследования. Современные требования к системам обработки и передачи речевых сигналов (РС), особенно при ограниченных ресурсах канала связи, требуют разработки эффективных методов многоскоростной обработки и кодирования. Классическая теорема дискретизации В.А. Котельникова [1] является основой цифровой обработки сигналов, однако её применимость ограничена стационарными сигналами, в то время как речь – нестационарный сигнал со спектральными характеристиками, динамически изменяющимися во времени. Это стимулирует развитие обобщённых подходов к дискретизации и представлению РС.

Практическая реализация теоремы В.А. Котельникова сопряжена с рядом ошибок: усечением ряда отсчётов [2–7], наложением спектров из-за нарушения условия ограниченности спектра [8–16], джиттером временных интервалов, а также ошибками округления и амплитудными погрешностями [7,8]. Эти недостатки особенно критичны для речи с резкими фонемными переходами и нестационарной интенсивностью. В таких условиях классическая дискретизация оказывается неадекватной, что обусловливает применение обобщённых методов, в частности — многоканального представления Хургина – Яковлева, реализующего принципы многоскоростной фильтрации на основе кратной децимации и интерполяции [9–11]. Оно использует банки фильтров для совместной обработки отсчётов сигнала и его производных, обеспечивая более точное восстановление нестационарных особенностей речи. Однако исходная версия представления не гарантирует точного восстановления из-за искажений фазового спектра синтезирующего фильтра, что требует его модификации путём коррекции фазочастотной характеристики [12–14].

К этому же классу методов многоскоростной обработки относится дискретное вейвлет-преобразование (ДВП), обладающее хорошей локализацией во времени и частоте, что делает его эффективным для анализа и сжатия речи [15–56]. Пороговая обработка вейвлет-коэффициентов применяется для подавления шумов и сжатия, однако выбор порогов часто основан на универсальных эмпирических правилах, не учитывающих статистику речи и условия передачи. Это приводит либо

к потере речевой информации, либо к недостаточному подавлению шумов. Решение возможно за счёт глобальной оптимизации, в частности — алгоритма роя частиц (Particle Swarm Optimization — PSO), который эффективен при настройке многомерных нелинейных систем и легко адаптируется к изменчивым условиям [57–70]. Вместе с тем, в научной литературе практически отсутствуют исследования, посвящённые совместному применению алгоритма роя частиц, обобщённой дискретизации на основе отсчётов сигнала и его производных (в рамках представления Хургина – Яковлева) и вейвлет-анализа РС.

Таким образом, актуальность настоящего исследования обусловлена следующими факторами:

– принципиальным несоответствием классической теории равномерной дискретизации динамической природе РС;

– отсутствием практических алгоритмов многоскоростной обработки РС, реализующих потенциал обобщённой многоканальной дискретизации на основе отсчётов сигнала и его производных;

– необходимостью адаптивной настройки порогов вейвлет-коэффициентов с учётом статистических и временных особенностей речи;

– потребностью в методах, обеспечивающих оптимальный компромисс между качеством восстановленной речи и объёмом передаваемых данных.

Все перечисленные обстоятельства подтверждают актуальность разработки алгоритмов многоскоростной обработки РС для радиотехнических систем передачи информации, основанных на комбинации обобщённой дискретизации с использованием представления Хургина – Яковлева, ДВП и адаптивной пороговой обработки с оптимизацией параметров посредством алгоритма роя частиц.

Степень проработанности темы исследования. Проблематика дискретизации сигналов, включая обобщённые схемы с использованием отсчётов производных и многоканальные методы восстановления, исследовалась широким кругом российских и зарубежных учёных. Основы классической теории дискретизации заложены В.А. Котельниковым [1], К.Э. Шенноном [2, 3], Э.Т. Уиттекером [4, 5] и Г. Найквистом [6]. Обобщение теоремы отсчётов на

многоканальные системы с использованием производных сигнала разработано в работах Я.И. Хургина, В.П. Яковлева [11–14], Дж. Папулиса [15], М. Веттерли [16], П.Л. Драготти [17], Т. Блу [18, 19] и др. Однако большинство существующих решений не учитывают фазовые искажения при восстановлении нестационарных РС и не ориентированы на условия ограниченной пропускной способности радиоканалов.

В области вейвлет-анализа фундаментальный вклад внесли С. Малла [29, 30], И. Добеши [31], Р.Р. Коифман [33], Д.Л. Доньо, А. Коэн [34] и М. Унсер [35], заложившие основы разрежённых представлений и пороговой обработки. В частности, методы подавления шумов и сжатия на основе универсальных и адаптивных порогов (например, VisuShrink, SureShrink) широко применяются, но редко адаптируются под статистику речи или условия передачи в реальных каналах. В российской научной школе значительные результаты в теории вейвлетов и их применении к анализу сигналов получены И.Я. Новиковым [38, 39], М.А. Скопиной [40], Е. А. Лебедевой [41, 42], Ю. А. Фарковым [43], В. В. Витязевым [98], Ю. С. Бехтиным и др [44–50].

Применение алгоритма роя частиц PSO для настройки параметров обработки сигналов исследовалось в работах Дж. Кеннеди [57], Р.К. Эберхарта [58], М. Клерка [61], Ю. Ши [59] и Х. Ху [60] и др [62–66]. В российской научной среде вопросы адаптации роевых алгоритмов к задачам обработки сигналов рассматривались В.В. Карпенко [67], И.А. Ходашинским и Е.С. Семёновым.

Несмотря на обширную научную базу, в литературе отсутствуют исследования, объединяющие в единую модель три ключевых направления:

– многоскоростную многоканальную дискретизацию РС по отсчётам и его производным;

– адаптивную пороговую обработку коэффициентов ДВП;

– оптимизацию параметров порогов методом роя частиц PSO.

Цель и задачи исследования. Целью диссертационного исследования является разработка алгоритмов многоскоростной обработки РС в радиотехнических системах передачи информации на основе представления Хургина – Яковлева и ДВП с оптимизацией параметров пороговой обработки.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Разработать математическую модель на основе представления Хургина – Яковлева для многоканальной дискретизации РС.

2. Исследовать возможности реализации методов первичного кодирования РС на основе представления Хургина – Яковлева.

3. Разработать метод адаптивной пороговой обработки вейвлет-коэффициентов на основе энергетического критерия с применением алгоритма роя частиц PSO и интегрировать его в комплексную систему кодирования.

4. Выполнить экспериментальную оценку эффективности разработанных алгоритмов на основе объективных и перцептивных метрик качества в сравнении с существующими методами кодирования речи.

Объект исследования. Алгоритмы многоскоростной обработки нестационарных РС в условиях ограниченной пропускной способности радиотехнических каналов передачи информации.

Предмет исследования. Алгоритмы многоскоростной обработки РС, основанные на представлении Хургина – Яковлева, ДВП с адаптивной пороговой обработкой коэффициентов и оптимизацией параметров порогов методом роя частиц PSO для применения в радиотехнических системах передачи информации.

Методы исследования. В данной диссертационной работе использовались аппарат математического анализа, теория цифровой обработки сигналов, методы вейвлет-анализа и глобальной оптимизации, а также методы компьютерного моделирования. Экспериментальная проверка эффективности предложенных алгоритмов осуществлялась на реальных и синтетических РС в условиях, имитирующих ограничения радиотехнических каналов передачи информации.

Научная новизна. В рамках диссертационной работы получены следующие новые научные результаты.

1. Разработан алгоритм дискретизации и первичного кодирования речевых сигналов на основе представления Хургина – Яковлева с многоканальной реализацией и многоуровневым разложением. (*Соответствует* ***п. 4 паспорта специальности 2.2.13*** *Разработка и исследование устройств генерирования,*

усиления, преобразования и синтеза радиосигналов, сигналов изображения и звука в радиотехнических системах различного назначения, включая системы телевидения. Создание эффективных методов их расчета и основ проектирования).

2. Предложен метод адаптивной настройки энергетических порогов дискретного вейвлет-преобразования для сжатия речевых сигналов на основе алгоритма роя частиц PSO. (*Соответствует* ***п. 9 паспорта специальности 2.2.13*** *Исследование и разработка телевизионных систем и устройств, включая матричные фотоприемники, телевизионные камеры, в том числе с IP-интерфейсами, цифровые кодеки, модемы и устройства передачи и воспроизведения изображений и звука, а также методов и алгоритмов модуляции, демодуляции, сжатия информации, помехоустойчивого кодирования и декодирования сигналов*).

3. Разработана интегрированная модель первичного кодирования речи, объединяющая обобщённую дискретизацию на основе представления Хургина – Яковлева, вейвлет-анализ и оптимизацию энергетических порогов методом роя частиц PSO. (*Соответствует* ***п. 6 паспорта специальности 2.2.13*** *Разработка и исследование методов и алгоритмов обработки сигналов и информации в радиотехнических устройствах и системах различного назначения, включая системы телевидения и передачи информации, при наличии помех с целью повышения помехоустойчивости*).

Основные положения, выносимые на защиту

1. Модификация синтезирующих фильтров для представления Хургина – Яковлева, обеспечивающая возможность многоскоростной обработки с точностью восстановления не менее 45 дБ для двухканальных систем и не менее 14 дБ для трёх- и четырёхканальных систем. (*Соответствует* ***п. 4 паспорта специальности 2.2.13*** *Разработка и исследование устройств генерирования, усиления, преобразования и синтеза радиосигналов, сигналов изображения и звука в радиотехнических системах различного назначения, включая системы телевидения. Создание эффективных методов их расчета и основ проектирования*).

2. Алгоритмы кодирования речевых сигналов на основе представления Хургина – Яковлева, повышающие качество восстановленной речи по сравнению с системой на основе теоремы В.А. Котельникова на величину до 2,2 балла по шкале MOS (Mean Opinion Score) в системах передачи речи различной скорости при вероятности битовых ошибок в канале до 5%. (*Соответствует* ***п. 6 паспорта специальности 2.2.13*** *Разработка и исследование методов и алгоритмов обработки сигналов и информации в радиотехнических устройствах и системах различного назначения, включая системы телевидения и передачи информации, при наличии помех с целью повышения помехоустойчивости*).

3. Алгоритм адаптивной настройки энергетических порогов дискретного вейвлет-преобразования с оптимизацией параметров алгоритмом роя частиц PSO, увеличивающий коэффициент сжатия передаваемой речи на величину до 14% по сравнению с классическими методами пороговой обработки при сохранении заданного уровня качества восстановленной речи. (*Соответствует* ***п. 9 паспорта специальности 2.2.13*** *Исследование и разработка телевизионных систем и устройств, включая матричные фотоприемники, телевизионные камеры, в том числе с IP-интерфейсами, цифровые кодеки, модемы и устройства передачи и воспроизведения изображений и звука, а также методов и алгоритмов модуляции, демодуляции, сжатия информации, помехоустойчивого кодирования и декодирования сигналов*).

Достоверность полученных результатов обусловливается использованием в процессе исследований адекватных математических моделей, корректным использованием математического аппарата, подтверждением полученных теоретических результатов имитационным моделированием на реальных речевых сигналах и сравнением полученных данных с результатами других авторов, а также с результатами экспериментальных исследований.

Практическая значимость. Разработанные алгоритмы и модели могут быть использованы при проектировании систем передачи и хранения речевой информации в условиях ограниченной пропускной способности каналов связи, включая радиотехнические, мобильные и спутниковые системы. Результаты работы

могут быть интегрированы в программно-аппаратные комплексы для цифровой обработки речи, а также использованы в учебном процессе при подготовке специалистов в области теории сигналов и систем связи.

Реализация и внедрение результатов работы. Результаты диссертационной работы внедрены:

1. В учебный процесс ФГБОУ ВО «Рязанский государственный радиотехнический университет им. В. Ф. Уткина»;

2. В ПАО «Мобильные ТелеСистемы», филиал г. Рязань;

3. В ООО «FPT Telecom», во Вьетнаме.

Личное участие автора. Автору на правах соавтора принадлежат постановки задач, разработка способов и алгоритмов их решения, отраженные в главах диссертации, а также компьютерное моделирование работы алгоритмов и анализ экспериментальных данных. Степень авторства подтверждается соответствующим заключением кафедры РУС РГРТУ.

Апробация работы. Результаты и положения диссертационной работы докладывались и обсуждались на 1 межрегиональной, 6 всероссийских и 9 международных научно-технических конференциях:

1. 6-я Межрегиональная научная сессия молодых исследователей «Промышленная революция 4.0: взгляд молодежи» (Тула, 2024 г.);

2. 28-я, 29-я, 30-я Всероссийская научно-техническая конференция студентов, молодых ученых и специалистов «Новые информационные технологии в научных исследованиях» (Рязань, 2023, 2024 и 2025 гг.);

3. Всероссийская научно-техническая конференция «Актуальные проблемы радиоэлектроники и телекоммуникаций» (Самара, 2025 г.);

4. 9-я, 10-я Всероссийская научно-техническая конференция «Актуальные проблемы современной науки и производства» (Рязань, 2024 и 2025 гг.);

5. 7-й, 8-й, 9-й, Международный научно-технический форум «Современные технологии в науке и образовании» (Рязань, 2024, 2025 и 2026 гг.);

6. 8-я Международная научно-практическая конференция «Актуальные проблемы и перспективы развития радиотехнических инфокоммуникационных

систем» «РАДИОИНФОКОМ-2024» (Москва, 2024 г.).

7. 8-я Международная научно-техническая конференция «Радиотехника, электроника и связь» «РЭиС-2025» (Омск, 2025 г.).

8. 11-я Международная заочная научно-техническая конференция «Информационные технологии. Радиоэлектроника. Телекоммуникации» «ITRT-2024» (Тольятти, 2024 г.).

9. 20-я и 21-я Международная научно-практическая конференция «Электронные средства и системы управления» (Томск, 2024 и 2025 гг.).

10. 31-я Международная научно-практическая конференция «Радиолокация, навигация, связь» (Воронеж, 2026 г.).

Публикации. По теме диссертации опубликовано 25 работ, в том числе 4 статьи в изданиях, входящих в список рекомендованных ВАК для опубликования результатов кандидатских диссертаций, 16 тезисов и материалов докладов на всероссийских и международных научно-технических конференциях, получено 5 свидетельств о государственной регистрации программ для ЭВМ.

Структура и объём диссертации. Диссертация состоит из введения, трех глав, заключения, приложения и списка литературы из 134 наименований и 4 приложений. Диссертации содержит 185 страницах, в том числе 150 основного текста, 11 таблиц и 41 рисунков.

ГЛАВА 1 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ АЛГОРИТМОВ МНОГОСКОРОСТНОЙ ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ НА ОСНОВЕ ПРЕДСТАВЛЕНИЯ ХУРГИНА – ЯКОВЛЕВА

1.1 Вводные замечания

Современные радиотехнические системы передачи речевой информации функционируют в условиях жёстких требований к эффективности использования полосы пропускания, вычислительных ресурсов и энергетических ограничений. Особенно это актуально для мобильных, спутниковых и аварийно-спасательных систем, где необходимо обеспечивать высокую разборчивость и естественность звучания речи при минимальном расходе канальных ресурсов. Традиционный подход к цифровому представлению речи основан на стандарте дискретизации с частотой 8 кГц, принятом для диапазона частот (300…3400 Гц) [1]. Этот подход теоретически обоснован теоремой о дискретизации В.А. Котельникова, но его практическая реализация неизбежно сопровождается рядом погрешностей.

Во-первых, точное восстановление сигнала требует бесконечного числа отсчётов, тогда как в реальных системах используется лишь конечное их количество, что приводит к ошибке усечения [7]. Во-вторых, речевые сигналы не являются строго ограниченными по спектру, поэтому при дискретизации возникает ошибка наложения спектров (алиасинг), даже при использовании антиалиасинговых фильтров [8]. В-третьих, нестабильность тактового генератора вызывает джиттер-ошибку, связанную с отклонением моментов взятия отсчётов от идеальных значений [9]. Кроме того, в процессе аналого-цифрового преобразования неизбежны ошибки округления и ошибки амплитуды, обусловленные конечной разрядностью аналого-цифрового преобразователя (АЦП) и неточностью измерения мгновенных значений сигнала [10]. Совокупность перечисленных факторов ограничивает точность представления и помехоустойчивость традиционных систем при фиксированной частоте дискретизации.

Таким образом, возникает фундаментальное противоречие: существующие методы дискретизации, основанные на теореме В.А. Котельникова, используют лишь мгновенные значения отсчётов сигнала и не учитывают нестационарность и

неравномерность информативности речи, что не позволяет снизить скорость передачи без потери качества; тогда как практика разработки низкоскоростных систем связи требует передавать РС с приемлемым качеством при существенно меньших затратах канального ресурса.

Перспективным путём разрешения указанного противоречия является представление Хургина – Яковлева [11] — математически строгое обобщение теоремы В.А. Котельникова, лежащее в основе алгоритмов многоскоростной обработки РС: снижение в K раз числа отсчётов сигнала компенсируется передачей (K – 1) производных, суммарный объём передаваемой информации сохраняется, а скорость передачи снижается.. В отличие от классического метода, данный подход использует в узлах дискретизации не только мгновенные значения сигнала, но и дополнительную информацию о характере его изменения — в виде производных различных порядков. Это особенно эффективно для РС, чья динамическая структура — с резкими переходами, взрывными согласными и быстро меняющимися формантами — делает информацию о скорости изменения сигнала высокоинформативной и позволяет точнее отслеживать временные вариации по сравнению с методами, основанными только на мгновенных значениях. Такое расширение исходного набора данных позволяет увеличить интервал между соседними узлами при сохранении условия точного восстановления, что открывает возможность снижения скорости передачи без потери качества.

Вместе с тем, практическая реализация представления Хургина – Яковлева сталкивается с принципиальной нерешённой проблемой, которая и составляет научную проблему настоящей диссертации: фазовые характеристики синтезирующих фильтров в исходной теоретической формулировке не определены однозначно, что приводит к фазовым искажениям восстановленного РС и исключает непосредственное практическое применение метода без его специализированной доработки.

Для решения указанной проблемы в настоящей работе предложена модификация, направленная на синтез конкретных амплитудно-частотных (АЧХ) и фазочастотных (ФЧХ) характеристик синтезирующих фильтров по критерию

минимума среднеквадратического отклонения. В результате разработан метод, обеспечивающий однозначное определение характеристик синтезирующих фильтров и гарантирующий точное восстановление РС с финитным спектром.

Предметом настоящей главы является разработка теоретических основ и методов синтеза характеристик синтезирующих фильтров для систем передачи и обработки РС на основе представления Хургина – Яковлева при числе каналов обработки $K = 2, 3, 4$, а также анализ точности разработанных систем.

В первой главе решаются следующие задачи:

1. Анализ спектральных и статистических характеристик РС, обосновывающий применимость и эффективность представления Хургина – Яковлева именно для данного класса сигналов.

2. Анализ теоретических основ обобщённого представления Хургина – Яковлева как математического обобщения теоремы В.А. Котельникова, обосновывающий возможность снижения скорости передачи без потери точности.

3. Модификация АЧХ и ФЧХ синтезирующих фильтров для двухканальной, трёхканальной и четырёхканальной систем по критерию минимума среднеквадратического отклонения.

4. Исследование методов вычисления производных дискретных сигналов и выбор оптимального алгоритма для реализации представления Хургина – Яковлева.

5. Экспериментальная оценка точности восстановления РС для систем с $K = 2,\ 3,\ 4$ каналами и $P = 1,\ 2,\ 3$ этапами разложения.

1.2 Особенности речевых сигналов и обоснование применимости представления Хургина – Яковлева

1.2.1 Статистические и спектральные характеристики речевых сигналов

Раздел посвящён анализу физических и статистических свойств РС с целью обоснования применимости и потенциальной эффективности представления Хургина – Яковлева для данного класса сигналов. В отличие от обзорного описания свойств речи, цель настоящего анализа состоит в доказательстве следующего тезиса: именно специфика РС делает информацию об их производных высокоинформативной, а значит, *K*-кратное увеличение интервала дискретизации

при K-канальной обработке достигается с минимальными потерями качества. Параллельно устанавливаются требования к системам передачи речи, определяющие критерии качества, при которых предлагаемый метод является конкурентоспособным по отношению к существующим стандартам кодирования.

С точки зрения теории сигналов речь является нестационарным случайным процессом с переменными во времени статистическими характеристиками. Ниже последовательно рассматриваются ключевые свойства РС с явным указанием следствий для представления Хургина – Яковлева.

1. Нестационарность и квазистационарность. На интервалах 10…30 мс РС может считаться квазистационарным, однако переходы между квазистационарными участками (взрывные согласные, фрикативы, аффрикаты) происходят чрезвычайно быстро за время порядка 5…15 мс и несут существенную информацию для разборчивости речи.

Следствием для представления Хургина – Яковлева является то, что именно на переходных участках первая производная сигнала достигает наибольших значений и содержит сконцентрированную информацию о характере изменения. Двухканальная система ($K = 2$) с передачей как отсчётов сигнала, так и его первой производной значительно точнее, чем одноканальная система В.А. Котельникова, воспроизводит быстрые изменения сигнала при том же числе отсчётов на единицу времени. Таким образом, нестационарный характер РС является прямым обоснованием эффективности метода при $K = 2$.

2. Спектральные характеристики и конечность полосы. Усреднённая спектральная плотность мощности РС демонстрирует спад порядка 5…6 дБ/октаву в диапазоне 1…4 кГц (рис. 1.1б). Согласно результатам Френча и Стейнберга [71], полоса 300…3400 Гц обеспечивает разборчивость слогов на уровне 95…98 %, что послужило основой для стандартизации телефонной полосы частот при частоте дискретизации 8 кГц.

Следствие для представления Хургина – Яковлева. Представление Хургина –Яковлева требует выполнения условия финитности спектра сигнала: $X(f) = 0$ при $|f| > F$. Интенсивный спад спектральной плотности речи за

пределами 4 кГц означает, что данное условие выполняется в практически значимом смысле с точностью, обеспечиваемой стандартными антиалиасинговыми фильтрами. Таким образом, спектральная ограниченность речи является необходимым условием применимости метода, и она надёжно обеспечивается в стандартных системах с полосой 0…4 кГц.

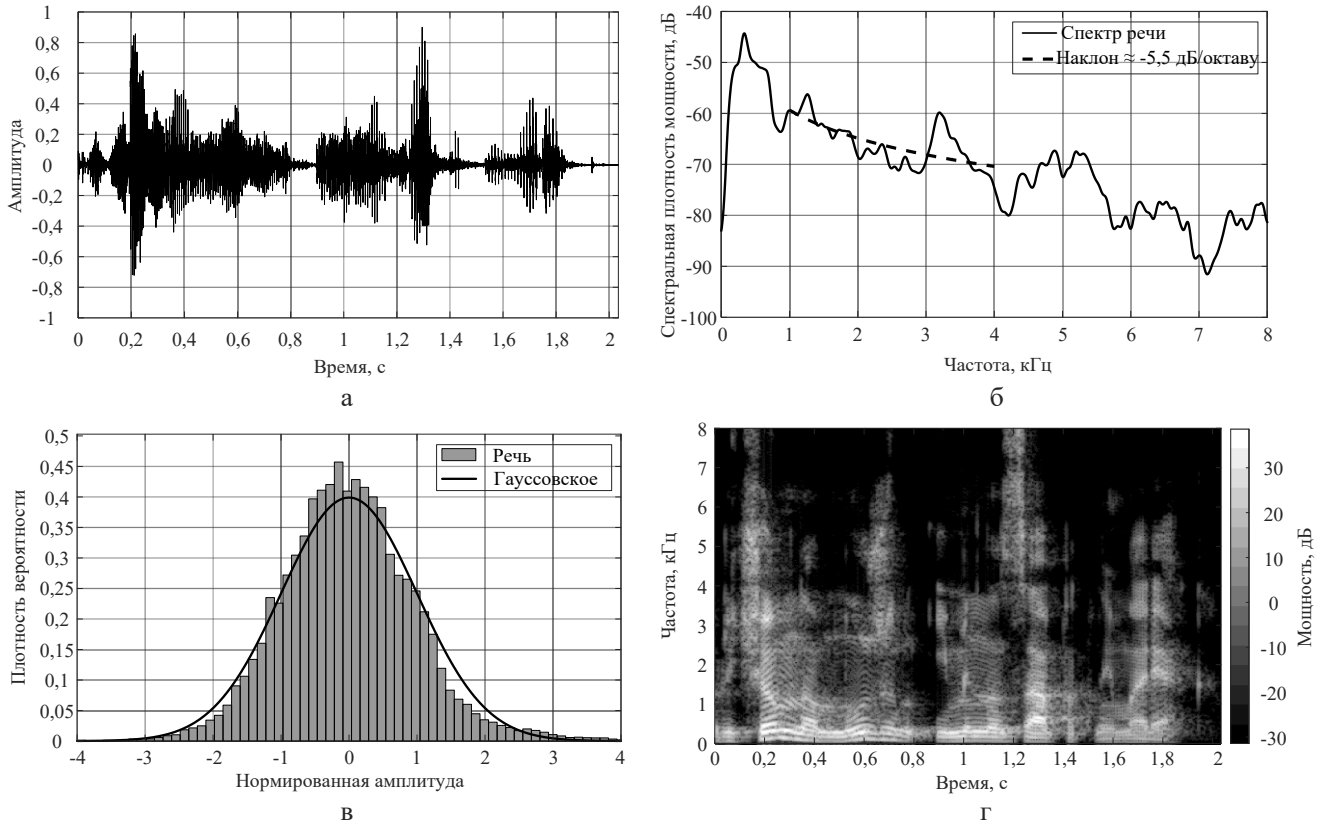


Рисунок 1.1 – Статистические характеристики РС:

(а) временная форма; (б) усредненный спектр мощности; (в) гистограмма амплитудного распределения; (г) частотно-временное представление

3. Амплитудные (негауссовские) характеристики. Гистограмма мгновенных значений РС существенно отличается от нормального закона (рис. 1.1в): коэффициент эксцесса составляет 1,5…2,5, а пик-фактор достигает 12…20 дБ. Равномерное квантование неоптимально для таких распределений; стандартным решением является логарифмическое квантование (А-закон, μ-закон).

Следствием для представления Хургина – Яковлева. Высокий пик-фактор означает, что в речевом сигнале кратковременные импульсные компоненты (взрывы, аффрикаты) несут значительную часть разборчивой информации. Первая производная чувствительна к подобным импульсным изменениям и позволяет

выделять их с высокой точностью при меньшем числе отсчётов. Это подтверждает, что двухканальная система $K=2$ наиболее эффективна именно для речи с её выраженными импульсными событиями.

4. Частотно-временная структура. Спектрограмма (рис. 1.1г) демонстрирует принципиальное различие между вокализованными сегментами (гармоническая структура, энергия в диапазоне 0…4 кГц) и невокализованными (широкополосный шумоподобный спектр, энергия в диапазоне 4…8 кГц). Эта структура используется в адаптивных методах кодирования для повышения эффективности сжатия.

Следствие для представления Хургина – Яковлева. Квазистационарность речи на интервалах 10…30 мс естественным образом соответствует покадровой реализации алгоритма Хургина – Яковлева: внутри каждого кадра сигнал характеризуется стабильными производными, что минимизирует погрешность их вычисления. Таким образом, временная структура речи не создаёт принципиальных препятствий для применения метода, а напротив, обеспечивает естественное разбиение на квазистационарные кадры.

5. Корреляционные свойства — ключевое обоснование двухканальной системы. Автокорреляционная функция речи характеризует статистическую связь между отсчётами на различных интервалах. Для вокализованных участков наблюдается периодическая структура с максимумами при задержках, кратных периоду основного тона (2,5…12,5 мс). Радиус корреляции составляет 5…10 отсчётов при частоте дискретизации 8 кГц [72].

Следствие для представления Хургина – Яковлева. Высокая корреляция соседних отсчётов означает, что первая производная сигнала несёт значительную, но при этом новую долю информации по сравнению с мгновенными значениями: она отражает не только текущую амплитуду, но и тенденцию изменения. Математически это означает, что вектор $\left(x\left(nKT\right),x'\left(nKT\right)\right)$ — совместные отсчёты сигнала и его производной — позволяет интерполировать сигнал на интервале $\left(nKT,\left(n+1\right)KT\right)$ с погрешностью, которая быстро убывает с ростом корреляции. В пределе при идеально коррелированном (постоянном) сигнале производная равна

нулю и не несёт дополнительной информации; при слабо коррелированном — производная приближается к белому шуму и теряет предсказательную ценность. Речевые сигналы занимают промежуточное оптимальное положение: достаточно высокая корреляция обеспечивает информативность производной, а нестационарность — её ненулевые значения. Именно это обстоятельство является фундаментальным обоснованием эффективности двухканальной системы $K = 2$ применительно к речи.

Аналогичная логика применима к трехканальным системам $K = 3$ и четырехканальным системам $K = 4$, где вторая и третья производные используются для восстановления сигнала на более широком интервале. Однако с ростом порядка производной её вычисление становится всё менее точным, что обусловливает нецелесообразность систем с $K > 4$ при данной частоте дискретизации.

1.2.2 Требования к качеству, скорости передачи и метрики оценки

Требования к системам передачи речи в радиотехнических каналах связи определяются компромиссом между качеством воспроизведения и эффективностью использования пропускной способности канала. Ниже систематизированы основные критерии, которые будут использованы при экспериментальной оценке предлагаемого в данной диссертации метода.

– Критерии качества и система объективных метрик. Качество передачи речи оценивается как субъективными, так и объективными методами. В данной диссертации используется комплекс из четырёх взаимодополняющих метрик, каждая из которых отражает различные аспекты качества.

Метрика 1. SNR (Signal-to-Noise Ratio) — отношение сигнал/шум (ОСШ) в децибелах, являющееся интегральной мерой общей энергии погрешности восстановления всего РС:

$$SNR = 10\lg\left(\sum_{n=0}^{N-1} x^2[n] \Big/ \sum_{n=0}^{N-1}\left(x[n] - \hat{x}[n]\right)^2\right), \tag{1.1}$$

где $x[n]$ – отсчеты исходного сигнала, $\hat{x}[n]$ – отсчеты восстановленного сигнала, N – общее количество отсчетов во всем анализируемом сигнале.

Метрика пригодна для интегральной оценки, но не учитывает

нестационарность речи: кратковременная высокая погрешность на тихих сегментах не вносит существенного вклада в глобальный SNR, хотя и заметна на слух. Для систем речевой связи минимально приемлемым считается уровень $\mathrm{SNR} \geq 13$ дБ [73].

Метрика 2. SegSNR (Segmental SNR) — сегментное ОСШ, характеризующее среднее качество восстановления сигнала по коротким перекрывающимся кадрам. Вычисляется по формуле:

$$SegSNR = \frac{1}{L}\sum_{l=0}^{L-1} 10\log_{10}\left(\sum_{n=0}^{M-1} x^2[n+lS] \Big/ \sum_{n=0}^{M-1}\left(x[n+lS]-\hat{x}[n+lS]\right)^2\right), \quad (1.2)$$

где $x[n]$ – отсчеты исходного сигнала, $\hat{x}[n]$ – отсчеты восстановленного сигнала, M – длина кадра (в отсчетах), S – шаг сдвига кадра (обычно 50% перекрытия), L – общее количество кадров в сигнале, l – индекс текущего кадра.

В отличие от глобальной метрики SNR, SegSNR учитывает нестационарную природу речи, поскольку усредняется по отдельным временным кадрам. Метрика является более чувствительной к кратковременным артефактам на тихих участках и более тесно коррелирует с субъективными оценками разборчивости. Согласно рекомендациям [71–73], приемлемым считается уровень $\mathrm{SegSNR} \geq 9{,}5$ дБ; высоким качеством считается $\mathrm{SegSNR} \geq 20$ дБ. Применение SegSNR наряду с глобальным SNR позволяет выявить случаи, когда алгоритм хорошо восстанавливает сигнал в среднем, но плохо справляется с конкретными фонетическими классами (взрывными, сибилянтами).

Метрика 3. MSE (Mean Squared Error) — среднеквадратическая погрешность, характеризует абсолютную энергию ошибки восстановления сигнала. Она определяется как среднее значение квадратов разностей между отсчетами исходного и восстановленного сигналов и вычисляется по следующей формуле:

$$MSE = \frac{1}{N}\sum_{n=0}^{N-1}\left(x[n]-\hat{x}[n]\right)^2, \quad (1.3)$$

где $x[n]$ – отсчеты исходного сигнала, $\hat{x}[n]$ – отсчеты восстановленного сигнала, N – общее количество отсчетов во всем анализируемом сигнале.

Метрика 4. Перцептивная метрика ViSQOL (Virtual Speech Quality Objective Listener) [74] оценивает воспринимаемое качество речи на шкале MOS (Mean

Opinion Score) в диапазоне 1,0…5,0 баллов. Метрика разработана в «Google Research» и обеспечивает коэффициент корреляции с субъективными оценками слушателей, превышающий значение 0,9. Пороговые значения: ViSQOL ≥ 4,0 – высокое качество (ITU-T категория «хорошо»); ViSQOL ≥ 3,5 – приемлемое качество для телефонии; ViSQOL < 3,0 – неприемлемое качество. В отличие от SNR, метрика ViSQOL учитывает психоакустические особенности восприятия: маскировку, характеристики критических полос слуха и нейросетевую модель внутреннего представления звука, и является предпочтительной метрикой для оценки новых методов кодирования [74].

Четыре перечисленные метрики (SNR, SegSNR, MSE, ViSQOL) в совокупности обеспечивают всестороннюю оценку качества восстановления РС: MSE отражает абсолютную точность алгоритма; SNR — интегральное соотношение сигнал/шум; SegSNR — качество на нестационарных участках; ViSQOL — воспринимаемое слушателем качество. Именно этот комплекс метрик используется в экспериментальных исследованиях настоящей диссертации.

– Соотношение качество-битрейт: Существующие методы кодирования обеспечивают различные компромиссы между скоростью передачи и качеством: 64 кбит/с (G.711) – ИКМ, MOS ≈ 4,1…4,2 балла, стандарт проводной цифровой телефонии [76]; 32 кбит/с (G.726) – АДИКМ, MOS ≈ 3,8…4,0 балла [77]; 13 кбит/с (GSM) – RPE-LTP, MOS ≈ 3,5…3,7 балла [78]; 8 кбит/с (G.729) – CS-ACELP, MOS ≈ 3,9…4,1 балла [79]; 2,4 кбит/с (MELP) – параметрическое кодирование, MOS ≈ 3,4…3,6 балла [81]. Зависимость качества от битрейта нелинейна и характеризуется областями насыщения и критическими точками. Для формантных синтезаторов критической является область 2…4 кбит/с, где небольшие изменения битрейта существенно влияют на качество.

– Помехоустойчивость: В радиоканалах с замираниями и помехами критична робастность кодека к битовым ошибкам. Методы прямого кодирования формы сигнала (ИКМ, АДИКМ) относительно устойчивы – ошибка в одном бите искажает 1…2 отсчета. Параметрические методы чувствительнее – ошибка в передаче параметров искажает целый фрейм длительностью 20…30 мс. Для повышения

помехоустойчивости применяются: Помехоустойчивое кодирование с неравномерной защитой (более важные биты защищены сильнее); Перемежение для борьбы с пакетными ошибками; Методы сокрытия ошибок на приемной стороне (интерполяция, замена на шум).

– Задержка в канале передачи: для дуплексных систем связи критична сквозная задержка. Согласно ITU-T G.114 [75], односторонняя задержка до 150 мс обеспечивает комфортный диалог; 150…400 мс допустима; более 400 мс воспринимается как полудуплексная связь. Алгоритмическая задержка кодека определяется длиной фрейма анализа и необходимостью упреждающего анализа. Методы прямого кодирования имеют минимальную задержку (<1 мс), параметрические методы — 20…60 мс.

– Полоса частот и частота дискретизации: Стандартная телефонная полоса 300…3400 Гц требует частоты дискретизации 8 кГц согласно теореме В.А. Котельникова. Широкополосная речь (50…7000 Гц) требует 16 кГц. Выбор полосы определяется компромиссом между качеством и требуемой пропускной способностью радиоканала. В условиях ограниченного спектра используется узкополосное представление.

– Вычислительная сложность: Реализация алгоритмов кодирования на цифровых сигнальных процессорах требует оценки вычислительной сложности. Измеряется в миллионах инструкций в секунду (MIPS). Простые методы требуют менее 1 MIPS, современные кодеки на основе линейного предсказания — 15…100 MIPS. Для мобильных устройств критично энергопотребление, пропорциональное вычислительной нагрузке.

Проведённый анализ показывает, что традиционные подходы к представлению РС в радиотехнических системах передачи основываются на равномерной дискретизации по теореме В.А. Котельникова. Это обеспечивает возможность точного восстановления, но не учитывает нестационарность и неравномерность информативности речи. Представление Хургина – Яковлева представляет перспективную альтернативу для повышения эффективности систем передачи. Его теоретические основы рассматриваются в следующем разделе.

1.3 Теорема отсчётов и обобщённое представление Хургина – Яковлева

1.3.1 Классическая теорема дискретизации В.А. Котельникова

Преобразование непрерывного сигнала в дискретную форму является фундаментальной операцией в радиотехнических системах передачи информации. Выбор метода дискретизации определяет как точность представления сигнала, так и эффективность использования пропускной способности канала связи. Рассмотрим основные теоретические положения, лежащие в основе дискретизации сигналов, начиная с классической теоремы отсчётов.

Теоретической основой цифровой обработки сигналов служит теорема отсчетов, независимо доказанная В.А. Котельниковым [1] и К. Шенноном [2]. Данная теорема устанавливает принципиальную возможность точного восстановления непрерывного сигнала по его дискретным отсчётам, что составляет теоретический фундамент для всех последующих рассуждений в данной главе.

Для сигнала $x(t)$, спектр которого строго ограничен диапазоном $(-\pi F;\pi F)$, точное восстановление непрерывного сигнала из дискретных отсчётов $x(nT)$, взятых с интервалом дискретизации $T=1/(2F)$, осуществляется по формуле:

$$x(t)=\sum_{n=-\infty}^{\infty}x(nT)\frac{\sin[\pi(t-nT)/T]}{\pi(t-nT)/T}, \tag{1.4}$$

где: $x(t)$ – непрерывный исходный сигнал; $x(nT)$ – дискретные отсчёты сигнала, взятые в моменты времени $t=nT$; n – целочисленный индекс отсчёта $(n=0,\pm1,\pm2,\ldots)$; T – интервал дискретизации; F – максимальная частота спектра сигнала. На основе формулы (1) реализована структурная схема, представленная на рисунке 1.2.

Дискретные сигналы $x(nT)$ формируются путем равномерной дискретизации непрерывного сигнала $x(t)$ с периодом T. Задача восстановления исходного непрерывного сигнала $x(t)$ из его дискретных отсчетов $x(nT)$ решается с использованием импульсной модуляции и фильтрации нижних частот.

Процесс восстановления включает следующие этапы:

– Дискретные значения сигнала $x(nT)$ умножаются на периодическую последовательность импульсов Дирака $\delta_T(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT)$, формируя модулированный сигнал.

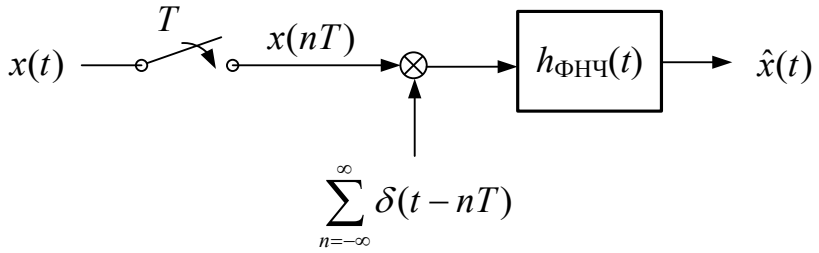


Рисунок 1.2 – Структурная схема системы дискретизации и восстановления сигнала на основе теоремы дискретизации В.А. Котельникова

– Полученный модулированный сигнал подается на вход аналогового фильтра нижних частот (ФНЧ), который выполняет интерполяцию между отсчетами и восстанавливает исходный непрерывный сигнал $x(t)$. В данном случае импульсная характеристика ФНЧ математически описывается выражением:

$$h_{\text{ФНЧ}}(t) = \frac{\sin(\pi t/T)}{\pi t/T}. \tag{1.5}$$

1.3.2 Обобщённое представление Хургина – Яковлева

Представление Хургина – Яковлева [11–14] является математически строгим обобщением теоремы В.А. Котельникова. Если доступны значения сигнала $x(t)$ и его первые $(K-1)$ производные, то для однозначного восстановления сигнала достаточно равномерно дискретизированных отсчётов, взятых в моменты времени с интервалом *KT*. Процесс восстановления исходного сигнала осуществляется по следующей формуле [11–14]:

$$x(t) = \sum_{k=0}^{K-1} \sum_{n=-\infty}^{\infty} \frac{1}{k!} x^{(k)}(nKT)(t - nKT)^k \left[\frac{\sin \pi(t/KT - n)}{\pi(t/KT - n)} \right]^K, \tag{1.6}$$

где T – период дискретизации сигнала, установленный согласно теореме В.А. Котельникова, не должен превышать $1/(2F)$; $x^{(k)}(nKT)$ – отсчетное значение *k-й* производной сигнала $x(t)$ взятой через равные временные интервалы *KT*.

Формулу (1.6) с математической точки зрения можно рассматривать как сумму *K* интерполированных рядов, *k-й* из которых имеет композиционную функцию:

$$h_{n,k}(t) = (t - nKT)^k \left[\frac{\sin \pi(t / KT - n)}{\pi(t / KT - n)} \right]^K . \qquad (1.7)$$

С точки зрения разработки системы передачи и обработки сигналов формулу (1.6) можно понимать следующим образом: для передачи сигнала $x(t)$, спектр которого сосредоточен в интервале $(-\pi/T; \pi/T)$, нужны отсчетные значения сигнала $x(nKT)$ и его первых $(K-1)$ производных $x'(nKT)$, $x''(nKT)$,..., $x^{(K-1)}(nKT)$ через равные временные интервалы *KT*. Эти отсчеты кодируются в виде коротких импульсов, площадь которых пропорциональна $x^{(k)}(nKT)/k!$, и передаются по *K* каналам соответственно.

На выходе приемника имеется *K* синтезирующих фильтров, имеющих импульсные переходные функции $h_{n,k}(t)$ $(k = 0,1,\ldots K-1)$, вычисляемые по формуле (1.7) в *K* различных каналах, и некоторое коммутационное устройство, подключающее последовательно выход приемника на вход этих фильтров таким образом, что на синтезирующий фильтр с номером *k* поступает импульс, соответствующий отсчету, взятому в момент $t_{n,k}$. Выходы всех фильтров объединяются в сумматоре, обеспечивая восстановленый сигнал $\hat{x}(t)$.

На основе метода передачи и обработки сигналов, предложенного Я.И. Хургиным и В.П. Яковлевым, описанного формулой (1.3), возможно разработать обобщенную структурную схему системы передачи и обработки сигналов, представленную на рисунке 1.3. Последовательность преобразований сигнала в данной схеме осуществляется следующим образом. На входе системы передачи исходный сигнал дискретизируется с частотой, определяемой теоремой В.А. Котельникова, в результате чего формируется дискретный сигнал $x(n)$. В *K*-канальной системе $(K-1)$ каналов содержат дифференциатор ($Д_k$) с целью получения отсчетов производных *k*-го порядка $(k = 1,\ldots,K-1)$. Дифференциатор ($Д_k$) функционирует согласно следующему алгоритму: сначала производится

быстрое преобразование Фурье (БПФ) отсчетов сигнала, затем отсчеты спектра умножаются на $(j2\pi f)^k$ и выполняется обратное быстрое преобразование Фурье [83]. После этого полученные дискретные отсчеты сигнала и его производных поступают в дециматор (↓K), где они прореживаются в K раз, затем квантуются в квантователе (КВ$_k$), передаются на кодер (КОД$_k$) и далее в канал связи (КС$_k$).

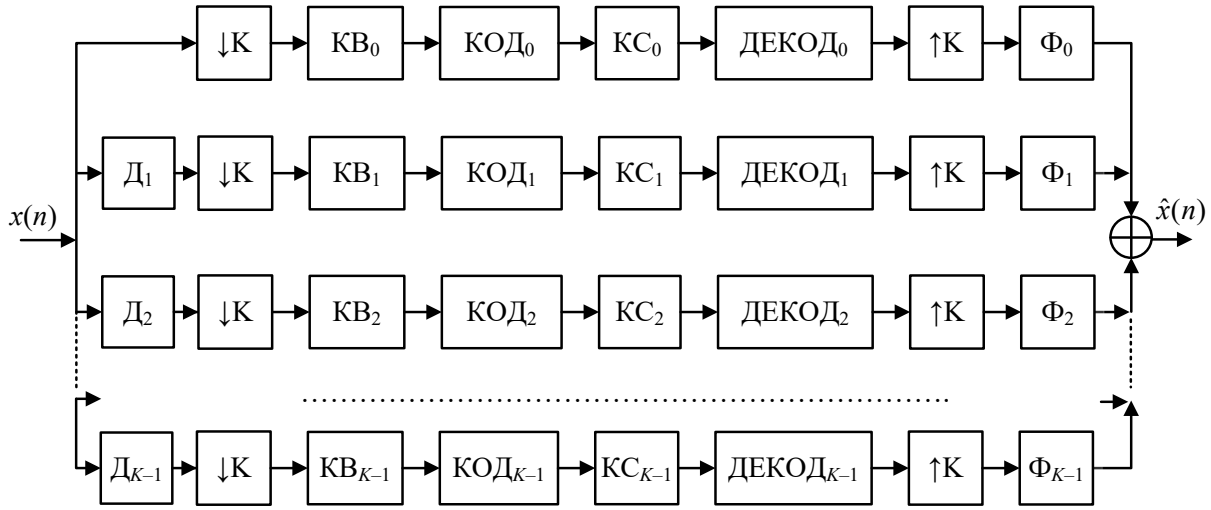


Рисунок 1.3 – Обобщенная структурная схема системы обработки и передачи сигналов на основе представления Хургина – Яковлева при K канальной реализации

После канала связи закодированные отсчеты поступают в K-канальный приемник. В декодере (ДЕКОД) происходит процедура декодирования согласно алгоритму первичного кодирования и затем передаются в интерполятор (↑K). На выходе интерполятора получаем отсчеты сигнала и отсчеты первой, второй, ... и $(K-1)$ производных. Эти отсчеты затем подаются на соответствующие синтезирующие фильтры (Ф$_k$), импульсные переходные функции $h_{n,k}(t)$ которых показаны ранее. Далее отсчеты поступают в сумматор (+), где происходит восстановление исходного сигнала $\hat{x}(n)$.

Рассмотрим альтернативный вариант реализации системы обработки и передачи сигналов на основе представления Хургина – Яковлева с использованием производных более высоких порядков – путем повторного применения представления Хургина – Яковлева на разных этапах. Структурная схема системы обработки и передачи сигналов на основе представления Хургина – Яковлева при реализации K каналов и P этапов представлена на рисунке 1.4.

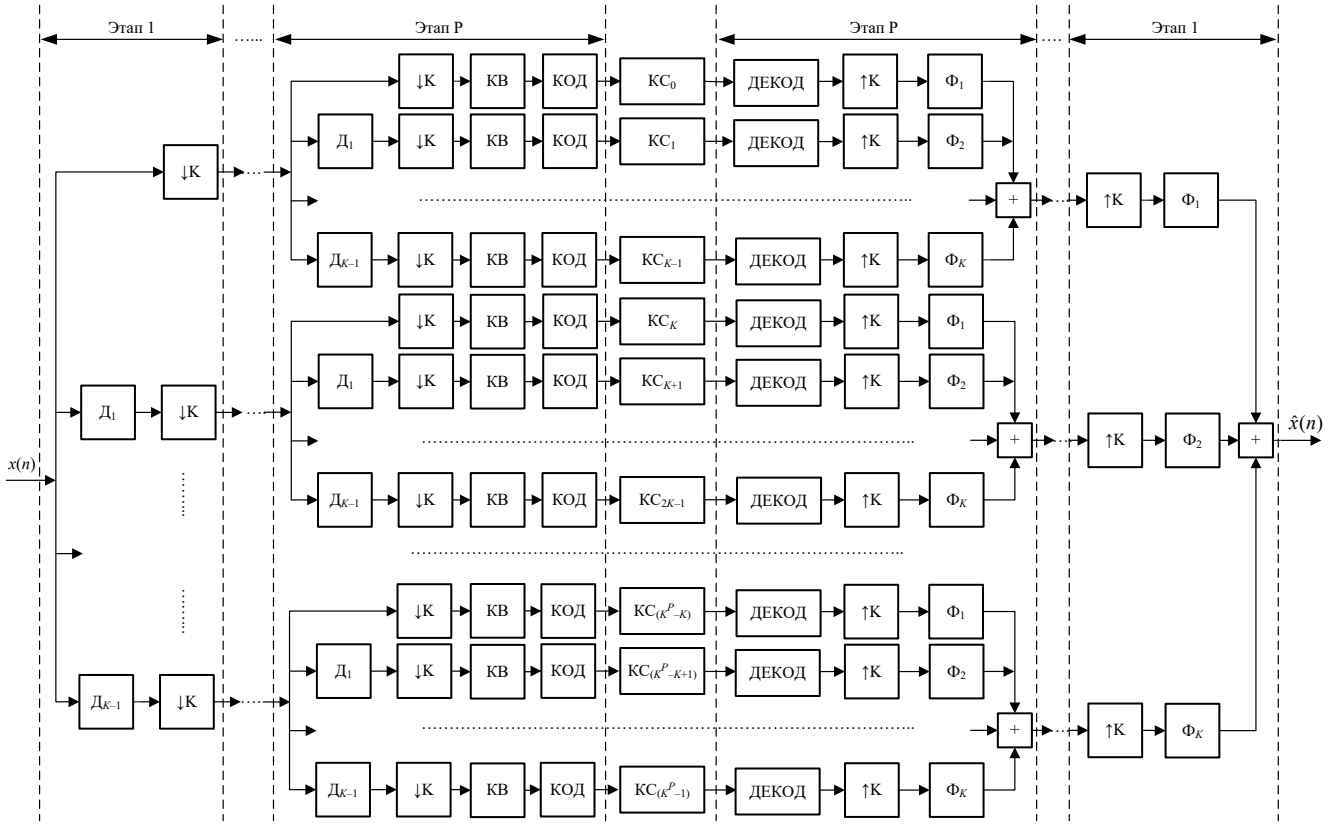


Рисунок 1.4 – Структурная схема система обработки и передачи сигналов на основе представления алгоритма Хургина – Яковлева при реализации K каналов и P этапов

Анализ структурной схемы (рисунок 1.4) показывает, что многоэтапное разложение сигнала в рамках представления Хургина – Яковлева при K каналах и P этапах приводит к формированию общего количества каналов обработки, равному K^P. Ключевым преимуществом данной архитектуры является сохранение принципа параллелизма вычислений. Однако реализация многоэтапной структуры сопряжена с ростом аппаратной сложности из-за увеличения количества узлов дифференцирования, децимации, интерполяции и синтеза. Тем не менее, высокая степень параллелизма позволяет эффективно распределить вычислительную нагрузку и компенсировать указанные затраты.

1.4 Разработка синтезирующих фильтров для многоканальных систем на основе представления Хургина – Яковлева

1.4.1 Двухканальная система на основе представления Хургина – Яковлева

Реализуя формулу (1.6) при различных значениях K, получаем системы

обработки сигналов на основе представления Хургина – Яковлева с различным количеством каналов. В рамках данного исследования рассматриваем случаи одноканальных, двухканальных, трехканальных и четырехканальных систем передачи при $K = 1 \ldots 4$.

Одноканальная система $(K = 1)$ соответствует теореме В.А. Котельникова, которая обеспечивает восстановление сигнала по дискретизированным отсчётам $x(nT)$ в фильтре низких частот.

При увеличении количества каналов до $K = 2 \ldots 4$, процедура восстановления сигнала в соответствии с представлением Хургина – Яковлева предполагает использование отсчетов самого сигнала $x(nKT)$ и его производных $x^{(k)}(nKT)$ до порядка $(K-1)$, в сочетании с синтезирующими фильтрами, обладающими импульсными передаточными функциями $h_{n,k}(t)$ по формуле (1.7). С их помощью возможно легко определить импульсные переходные характеристики (ИПХ), АЧХ и ФЧХ синтезирующих фильтров. Ключевой проблемой практической реализации представления Хургина – Яковлева является неопределённость ФЧХ синтезирующих фильтров, заданных базисными функциями (1.4). В рамках данного исследования разработан метод синтеза АЧХ и ФЧХ синтезирующих фильтров по критерию минимума среднеквадратического отклонения (СКО) для систем с $K = 2, 3, 4$ каналами.

При $K = 2$ исходный сигнал представляется как отсчеты самого сигнала и его первой производной. Это позволяет точно восстановить сигнал с финитным спектром, полностью сосредоточенным в интервале $(-\pi/T; \pi/T)$, по отсчётам, взятым с периодом $2T$, используя следующую формулу:

$$x(t) = \sum_{n=-\infty}^{\infty} x(n2T)\left[\frac{\sin\pi(t/2T-n)}{\pi(t/2T-n)}\right]^2 + \sum_{n=-\infty}^{\infty} x'(n2T)(t-n2T)\left[\frac{\sin\pi(t/2T-n)}{\pi(t/2T-n)}\right]^2. \quad (1.8)$$

Существенное отличие от классической теоремы В.А. Котельникова заключается в том, что объём информации о сигнале на единицу времени определяется не только значениями отсчётов, но и значениями их первой производной [11]. Метод передачи и обработки сигнала, основанный на формуле (1.8), представлен на рисунке 1.5.

Исходный сигнал подвергается дискретизации в соответствии с теоремой В.А. Котельникова. Дискретизированный сигнал поступает на дифференциатор ($Д_1$), формирующий отсчеты первой производной. Полученные дискретные последовательности сигнала $x(n)$ и его производной $x'(n)$ подаются в дециматор ($\downarrow 2$), который реализует операцию прореживания – отбрасывает каждый второй отсчет [14]. В результате формируются две прореженные последовательности — отсчётов сигнала $x(2n)$ и его производной $x'(2n)$, — которые далее подвергаются квантованию, кодированию и передаются по каналу связи.

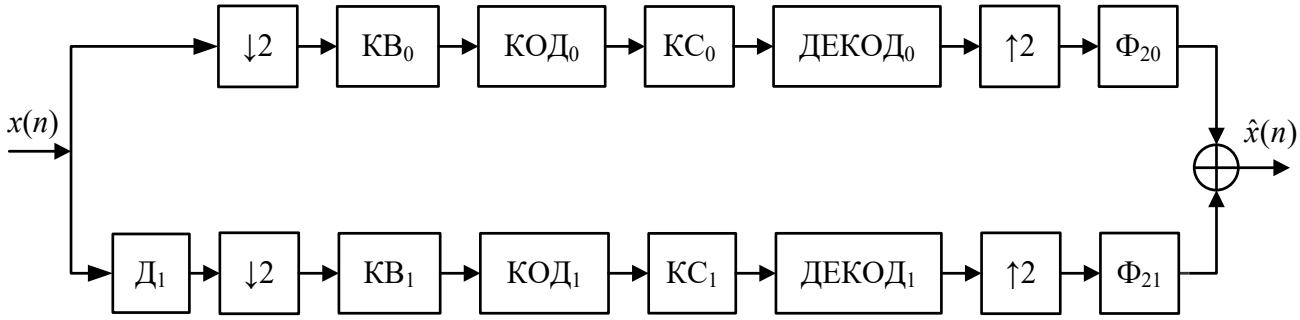


Рисунок 1.5 – Структурная схема системы обработки и передачи сигналов на основе представления Хургина – Яковлева при двухканальной реализации

Закодированные последовательности поступают на двухканальный приёмник. После декодирования выполняется двукратная интерполяция каждой из последовательностей: между соседними отсчётами вставляются нулевые значения. Полученные интерполированные последовательности подаются на синтезирующие фильтры $Ф_{20}$ и $Ф_{21}$, выходы которых суммируются, формируя на выходе восстановленный сигнал. Синтезирующие фильтры с импульсными переходными функциями, задаваемыми следующими формулами:

– для отсчетов сигнала:

$$h_{20}(t) = \left[\frac{\sin \pi \left(t / 2T - n \right)}{\pi \left(t / 2T - n \right)} \right]^2, \tag{1.9}$$

– для отсчетов первой производной сигнала:

$$h_{21}(t) = (t - 2nT) \left[\frac{\sin \pi \left(t / 2T - n \right)}{\pi \left(t / 2T - n \right)} \right]^2. \tag{1.10}$$

Исходные АЧХ и ФЧХ синтезирующих фильтров показаны на рисунке 1.6.

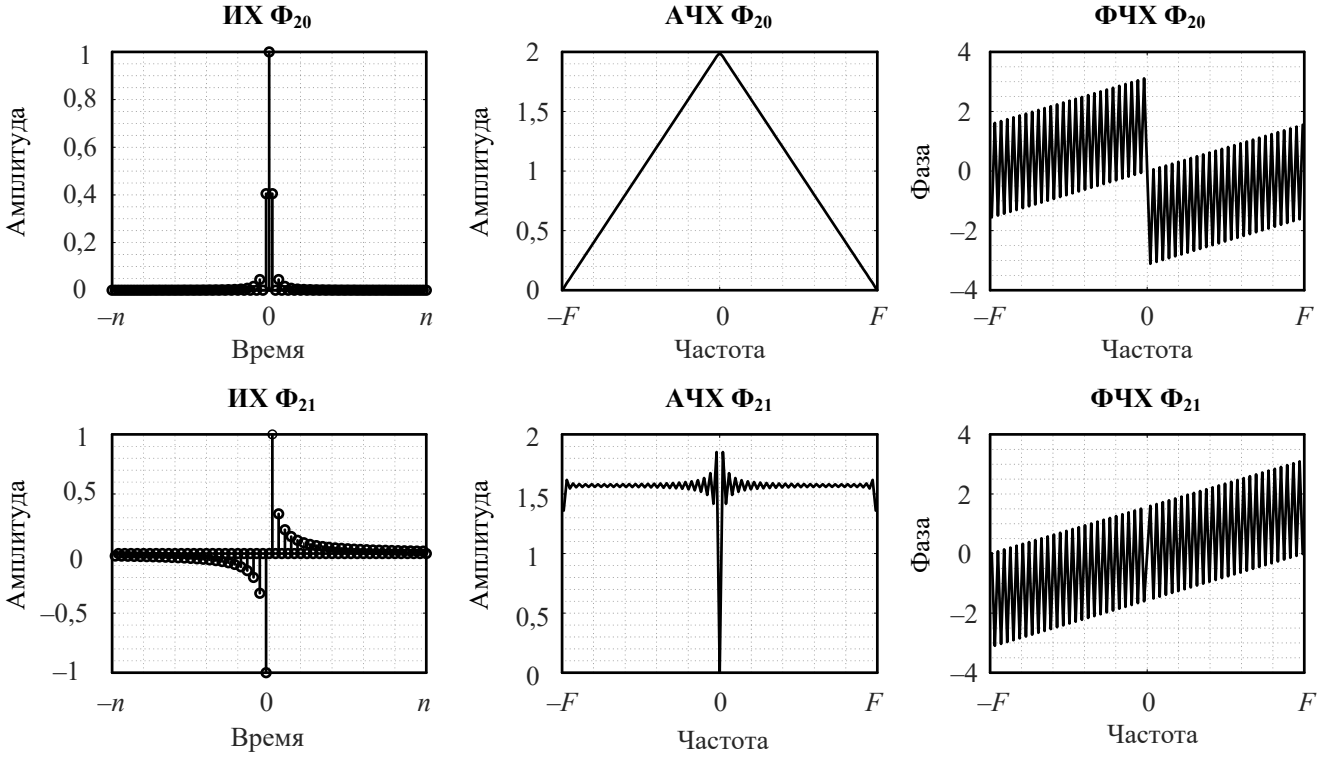


Рисунок 1.6 – АЧХ и ФЧХ синтезирующих фильтров
представления Хургина – Яковлева при двухканальной реализации

ФЧХ упомянутых выше синтезирующих фильтров Φ_{20} и Φ_{21} неопределены, что приводит к ошибкам при восстановлении сигнала на основе представления Хургина – Яковлева, ограничивая его применимость в практических задачах [3, 7]. Для решения данной проблемы предложена модификация алгоритма, направленная на уменьшение погрешности восстановления сигнала за счет коррекции ФЧХ, обеспечивающих более точное восстановление сигнала в системе связи и использования методов аппроксимации для вывода универсального уравнения АЧХ синтезирующих фильтров.

АЧХ $A(f)$ и ФЧХ $\phi(f)$ синтезирующих фильтров, полученные после модификации, имеют вид [3, 7]:

Для синтезирующего фильтра Φ_{20} (канал отсчётов сигнала):

$$A_{20}(f) = 2(1 - f / F),\ \phi_{20}(f) = 0. \tag{1.11}$$

Для синтезирующего фильтра Φ_{21} (канал отсчётов первой производной):

$$A_{21}(f) = \begin{cases} 0, \text{ при } f = 0; F, \\ 2/\pi, \text{ при } f \neq 0; F, \end{cases},\ \phi_{21}(f) = \begin{cases} 0, \text{ при } f = 0; F, \\ -\pi/2, \text{ при } f \neq 0; F. \end{cases} \tag{1.12}$$

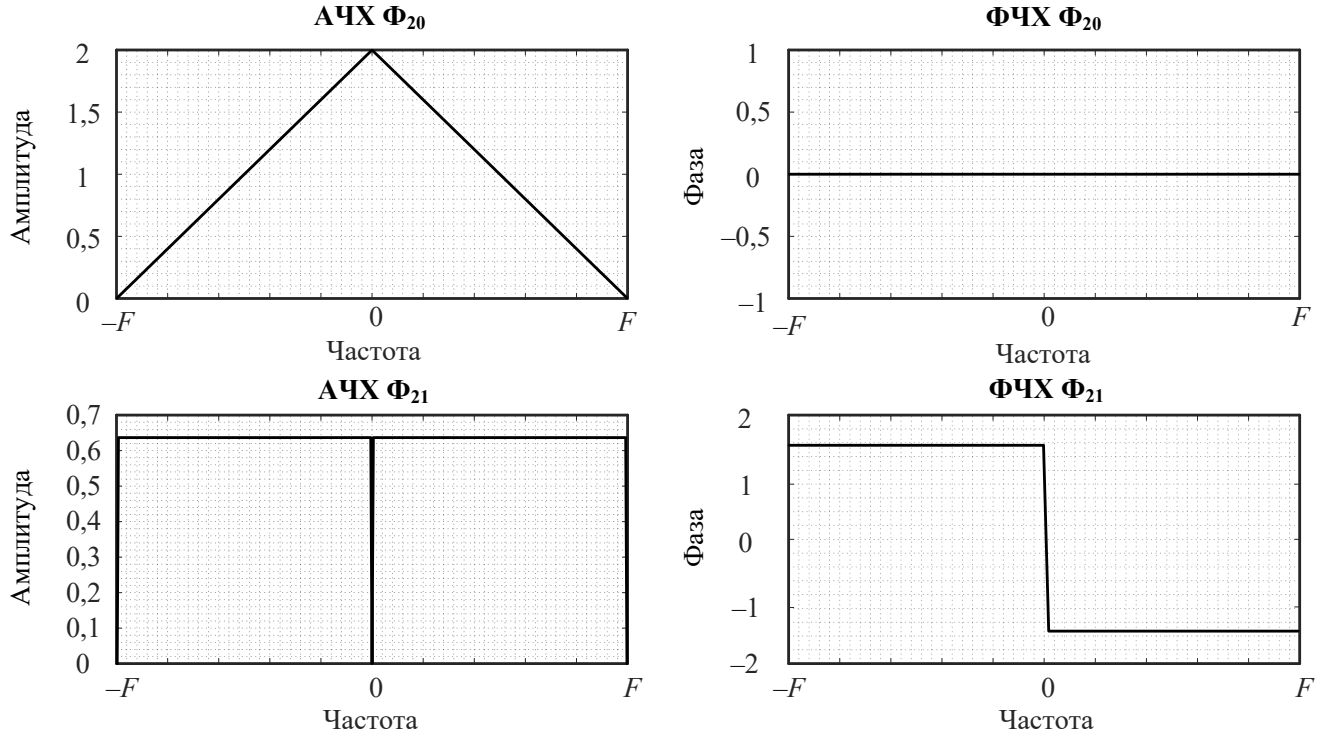


Рисунок 1.7 – АЧХ и ФЧХ синтезирующих фильтров
при двухканальной реализации после модификации

Покажем, что синтезированные характеристики (1.11), (1.12) обеспечивают идеальное восстановление сигнала $x(t)$ в диапазоне $(0;F)$. Пусть $X(f)$ — спектр исходного сигнала, $X(f)=0$ при $|f|>F$, частота дискретизации $F_{Д}=2F$ (период $T=1/(2F)$).

После децимации в два раза с периодом $2T=1/F$ и последующей двукратной интерполяции спектры в канале сигнала и канале первой производной принимают вид:

$$\tilde{X}_0(f)=F\left[X(f)+X(f-F)\right], \tag{1.13}$$

$$\tilde{X}_1(f)=F\left[j2\pi f.X(f)+j2\pi(f-F).X(f-F)\right]. \tag{1.14}$$

Слагаемое $X(f-F)$ в каждом из выражений (1.13), (1.14) представляет собой алиасинговую составляющую, неизбежно возникающую при децимации в два раза. Восстановленный сигнал на выходе сумматора имеет спектр:

$$\hat{X}(f)=H_{20}(f)\tilde{X}_0(f)+H_{21}(f)\tilde{X}_1(f). \tag{1.15}$$

Подставляя (1.13), (1.14) в (1.15) и группируя слагаемые при $X(f)$ и

$X(f-F)$, условие $\hat{X}(f)\equiv X(f)$ для произвольного сигнала равносильно выполнению двух соотношений при каждом $f\in(0;F)$:

$$F\left[H_{20}(f)+j2\pi f.H_{21}(f)\right]=1, \tag{1.16}$$

$$F\left[H_{20}(f)+j2\pi(f-F).H_{21}(f)\right]=0. \tag{1.17}$$

Уравнение (1.16) является условием верного воспроизведения сигнала, уравнение (1.17) — условием полного подавления алиасинговой составляющей. Вычитая (1.17) из (1.16), получаем:

$$F.j2\pi F.H_{21}(f)=1\Rightarrow H_{21}(f)=\frac{1}{j2\pi F^2}. \tag{1.18}$$

Подставляя выражение (1.18) в (1.17):

$$H_{20}(f)=-j2\pi(f-F).H_{21}(f)=\frac{F-f}{F^2}. \tag{1.19}$$

Таким образом, система (1.16), (1.17) имеет единственное решение, поскольку определитель матрицы системы $\Delta=-j2\pi F^2\neq 0$. Выразим синтезированные характеристики (1.11), (1.12) в комплексной форме при нормировке $F=1/2$:

$$H_{20}(f)=A_{20}(f)e^{j\phi_{20}(f)}=2(1-f/F)=2(F-f)/F, \tag{1.20}$$

$$H_{21}(f)=A_{21}(f)e^{j\phi_{21}(f)}=2/\pi.e^{-j\pi/2}=-2j/\pi. \tag{1.21}$$

Проверим выполнение условий (1.16), (1.17). Подставляя (1.20), (1.21) в левую часть (1.16):

$$F\left[\frac{2(F-f)}{F}+j2\pi f.\left(-\frac{2j}{\pi}\right)\right]=F\left[\frac{2(F-f)}{F}+j4f\right]=2(F-f)+4fF.$$

При $F=1/2$: $2(1/2-f)+4f.1/2=1-2f+2f=1$. Условие (1.16) выполнено.

Подставляя (1.20), (1.21) в левую часть (1.17):

$$F\left[\frac{2(F-f)}{F}+j2\pi(f-F).\left(-\frac{2j}{\pi}\right)\right]=2(F-f)+4F(f-F).$$

При $F=1/2$: $2(1/2-f)+4.1/2.(f-1/2)=(1-2f)+(2f-1)=0$. Условие (1.16) выполнено.

Следовательно, $\hat{X}(f) \equiv X(f)$ для всех $f \in (0;F)$. В силу эрмитовой симметрии спектра вещественного сигнала это равенство распространяется на весь диапазон $|f| < F$. Тем самым доказано, что синтезированные АЧХ и ФЧХ (1.11), (1.12) обеспечивают точное восстановление РС на выходе двухканальной системы на основе представления Хургина – Яковлева, полностью устраняя погрешность фазовой характеристики в рабочем диапазоне $(0;F)$.

1.4.2 Трёхканальная система на основе представления Хургина – Яковлева

При реализации системы передачи и обработки сигналов с тремя каналами, когда осуществляется передача отсчетов сигнала и его первых двух производных $(K=3)$, формула (1.6) принимает следующий вид:

$$x(t) = \sum_{n=-\infty}^{\infty} x(n3T)\left[\frac{\sin\pi(t/3T-n)}{\pi(t/3T-n)}\right]^3 + \sum_{n=-\infty}^{\infty} x'(n3T)(t-n3T)\left[\frac{\sin\pi(t/3T-n)}{\pi(t/3T-n)}\right]^3 + \\ + \sum_{n=-\infty}^{\infty} \frac{1}{2} x''(n3T)(t-n3T)^2\left[\frac{\sin\pi(t/3T-n)}{\pi(t/3T-n)}\right]^3. \quad (1.22)$$

Методика обработки сигнала, использующая формулу (1.22), выглядит следующим образом.

Передаваемый сигнал $x(t)$ имеет финитный спектр, сосредоточенный в диапазоне $(-\pi/T;\pi/T)$. Через интервалы времени длительности $3T$ выбираются отсчетные значения сигнала $x(3nT)$, его первой $x'(3nT)$ и второй производной $x''(3nT)$. Эти величины кодируются в виде коротких импульсов, амплитуды которых пропорциональны соответствующим величинам. После этого на приемной стороне они поступают на синтезирующие фильтры $(\Phi_{30},\Phi_{31},\Phi_{32})$ с импульсными переходными функциями, описываемыми следующими формулами:

– для отсчетов сигнала:

$$h_{30}(t) = \left[\frac{\sin\pi(t/3T-n)}{\pi(t/3T-n)}\right]^3. \quad (1.23)$$

– для отсчетов первой производной сигнала:

$$h_{31}(t) = (t - n3T)\left[\frac{\sin\pi(t/3T - n)}{\pi(t/3T - n)}\right]^3. \quad (1.24)$$

– для отсчетов второй производной сигнала:

$$h_{32}(t) = (t - n3T)^2\left[\frac{\sin\pi(t/3T - n)}{\pi(t/3T - n)}\right]^3. \quad (1.25)$$

АЧХ и ФЧХ синтезирующих фильтров, полученных для каналов прореженного сигнала и двух его первых производных, приведены на рисунке 1.8.

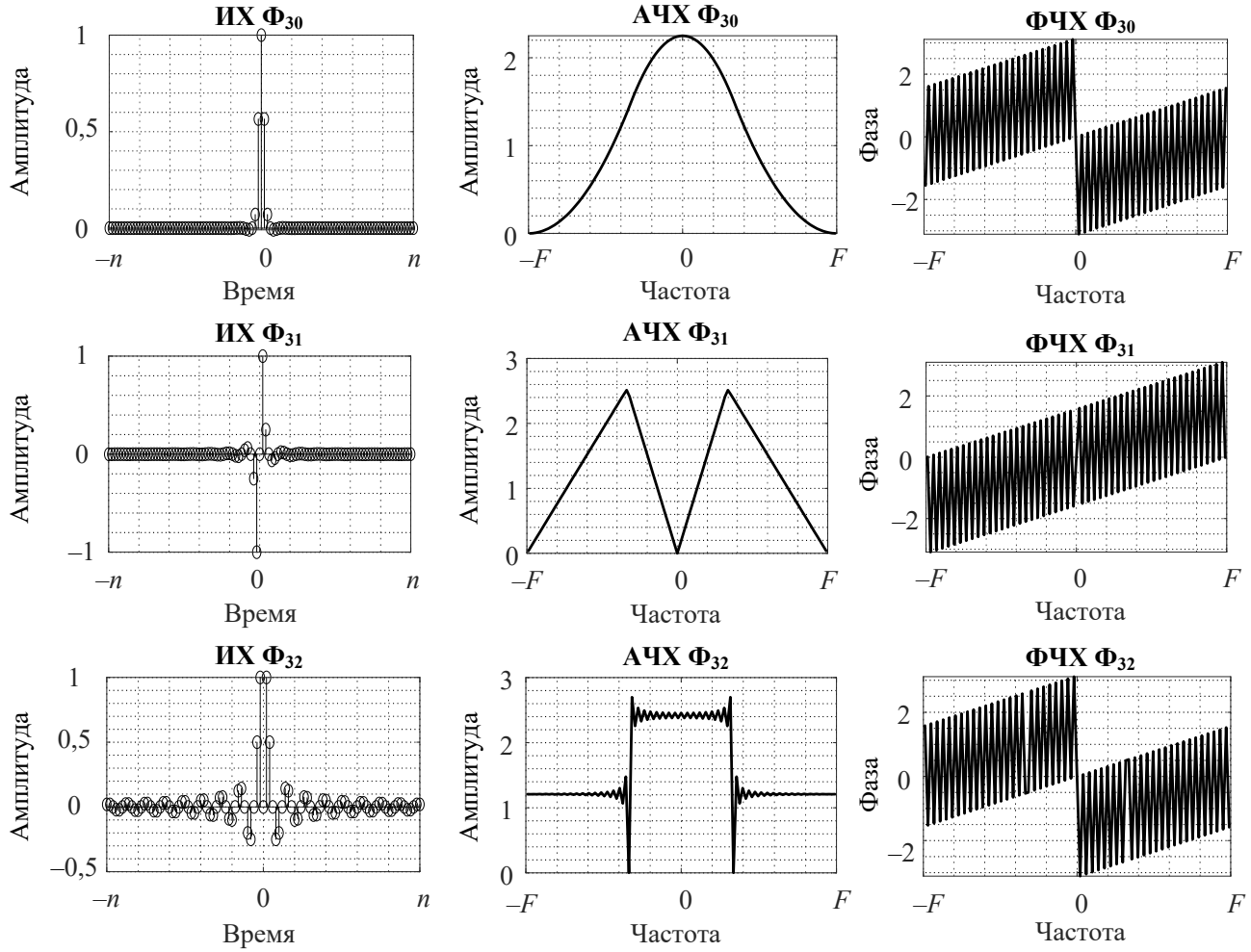


Рисунок 1.8 – АЧХ и ФЧХ синтезирующих фильтров
представления Хургина – Яковлева при трёхканальной реализации

Анализ рисунка 1.8 показывает, что ФЧХ синтезирующих фильтров не обеспечивают требуемой точности восстановления сигнала. Для устранения данного недостатка предлагается усовершенствовать подход к реализации представления Хургина – Яковлева путем коррекции ФЧХ. С этой целью для синтеза АЧХ и формирования универсального выражения применяются методы аппроксимации, позволяющие рассчитать желаемый частотный отклик с

использованием полиномиальной и кусочно-линейной аппроксимации. Полученные АЧХ $A(f)$ и ФЧХ $\phi(f)$ синтезирующих фильтров для трехканальной реализации представления Хургина – Яковлева представлены на рисунке 1.9.

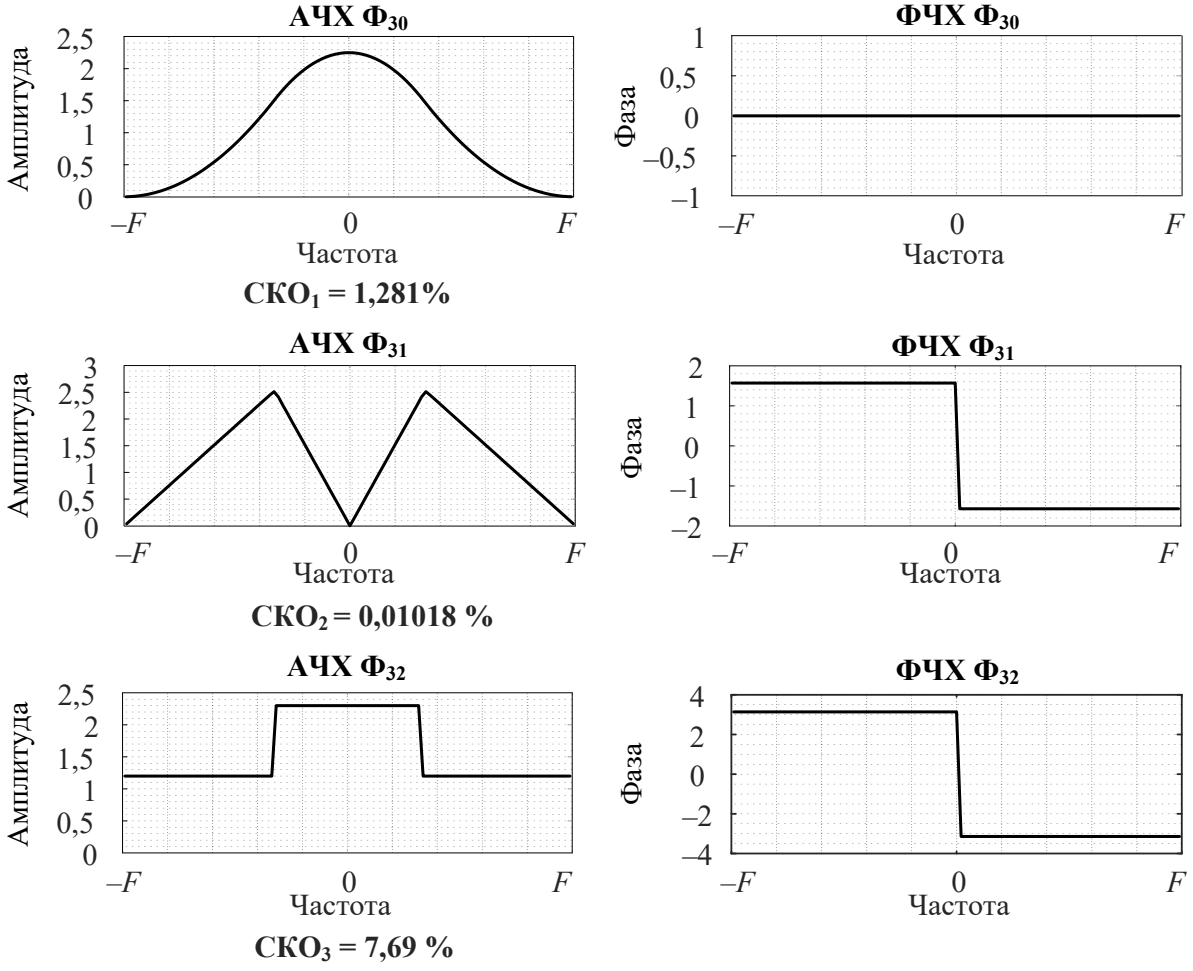


Рисунок 1.9 – АЧХ и ФЧХ синтезирующих фильтров
при трёхканальной реализации после модификации

АЧХ $A(f)$ и ФЧХ $\phi(f)$ синтезирующих фильтров в трехканальной системе $(K=3)$ после модификации имеют следующий вид [105]:

$$A_{30}(f)=2,22+1,25(f/F)-16,27(f/F)^2+20,45(f/F)^3-7,67(f/F)^4,\quad \phi_{30}(f)=0. \tag{1.26}$$

$$A_{31}(f)=\begin{cases} 4,26(f/F) \text{ при } f\le 0,36F,\\ -2,15(f/F)+2,15 \text{ при } 0,36F<f\le F.\end{cases}$$

$$\phi_{31}(f)=\begin{cases} -\pi/2 \text{ при } f\le F,\\ \pi/2 \text{ при } F<f\le 2F.\end{cases} \tag{1.27}$$

$$A_{32}(f)=\begin{cases} 1{,}37 \text{ при } f \le 0{,}36F, \\ 0{,}69 \text{ при } 0{,}36F < f \le F. \end{cases}$$
$$\phi_{32}(f)=\begin{cases} -\pi \text{ при } f \le F, \\ \pi \text{ при } F < f \le 2F. \end{cases} \qquad (1.28)$$

Структурная схема системы обработки и передачи сигналов на основе представления Хургина – Яковлева имеет вид, представленный на рисунке 1.10.

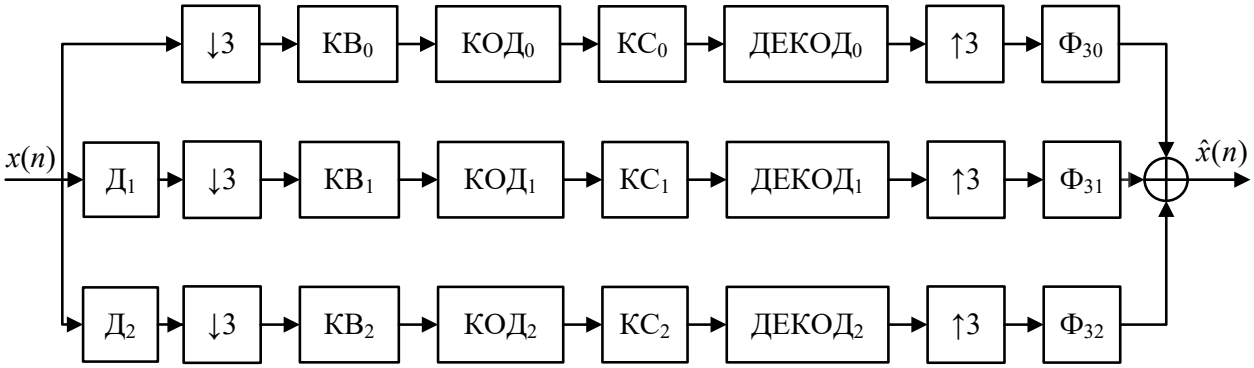


Рисунок 1.10 – Структурная схема системы обработки и передачи сигналов на основе представления Хургина – Яковлева при трёхканальной реализации

Первоначально происходит дискретизация исходного сигнала с частотой, определенной по теореме В.А. Котельникова $x(n)$. После этого, дискретизированный сигнал поступает в дифференциатор (Д$_k$) $\left(k=1,2\right)$, на выходе которого формируются отсчеты первой $x'(n)$ и второй $x''(n)$ производной сигнала. Следующим этапом является процесс децимации (↓3): из группы трёх отсчетов отбрасывается каждый второй и третий отсчеты как исходного сигнала, так и его первой и второй производных. В итоге мы получаем прореженные в 3 раза отсчеты сигнала и его отсчеты первой и второй производных.

Полученные отсчёты далее поступают на блоки квантования (КВ) и кодирования (КОД), а затем передаются в канал связи (КС). После прохождения канала связи, закодированные отсчеты поступают в трёхканальный приёмник. В декодере (ДЕКОД) происходит их декодирование и последующая передача на интерполятор (↑3), осуществляющий вставку двух нулевых отсчетов между исходными отсчетами сигнала. На выходе интерполятора получаем отсчеты сигнала и его отсчеты первой и второй производной, соответственно. Затем эти

отсчеты поступают на соответствующие синтезирующие фильтры $(\Phi_{30},\Phi_{31},\Phi_{32})$, после чего поступают на сумматор (+), на выходе которого восстанавливается исходный РС $\hat{x}(n)$.

1.4.3 Четырёхканальная система на основе представления Хургина – Яковлева

Переход к четырёхканальной системе (K = 4) обусловлен необходимостью дальнейшего снижения числа передаваемых отсчётов при сохранении приемлемого качества восстановления РС. При K = 4 передаются значения сигнала и его производных первого, второго и третьего порядков, что обеспечивает четырёхкратную децимацию по сравнению с классической теоремой В.А. Котельникова. Формула (1.6) принимает следующий вид:

$$x(t)=\sum_{n=-\infty}^{\infty}x(n4T)\left[\frac{\sin\pi(t/4T-n)}{\pi(t/4T-n)}\right]^4+\sum_{n=-\infty}^{\infty}x'(n4T)(t-n4T)\left[\frac{\sin\pi(t/4T-n)}{\pi(t/4T-n)}\right]^4+$$
$$+\sum_{n=-\infty}^{\infty}\frac{1}{2}x''(n4T)(t-n4T)^2\left[\frac{\sin\pi(t/4T-n)}{\pi(t/4T-n)}\right]^4+$$
$$+\sum_{n=-\infty}^{\infty}\frac{1}{6}x'''(n4T)(t-n4T)^3\left[\frac{\sin\pi(t/4T-n)}{\pi(t/4T-n)}\right]^4. \quad (1.29)$$

На стороне приемника сигналы поступают на синтезирующие фильтры (Φ_{40}, Φ_{41}, Φ_{42}, Φ_{43}), импульсные переходные функции которых определяются следующими выражениями:

– для отсчетов сигнала: $h_{40}(t)=\left[\frac{\sin\pi(t/4T-n)}{\pi(t/4T-n)}\right]^4$. (1.30)

– для отсчетов первой производной сигнала: $h_{41}(t)=(t-n4T)\left[\frac{\sin\pi(t/4T-n)}{\pi(t/4T-n)}\right]^4$. (1.31)

– для отсчетов второй производной сигнала:

$$h_{42}(t)=(t-n4T)^2\left[\frac{\sin\pi(t/4T-n)}{\pi(t/4T-n)}\right]^4. \quad (1.32)$$

– для отсчетов третьей производной сигнала:

$$h_{43}(t)=(t-n4T)^3\left[\frac{\sin\pi(t/4T-n)}{\pi(t/4T-n)}\right]^4. \quad (1.33)$$

АЧХ и ФЧХ синтезирующих фильтров, полученных для каналов прореженного сигнала и трех его первых производных, приведены на рисунке 1.11.

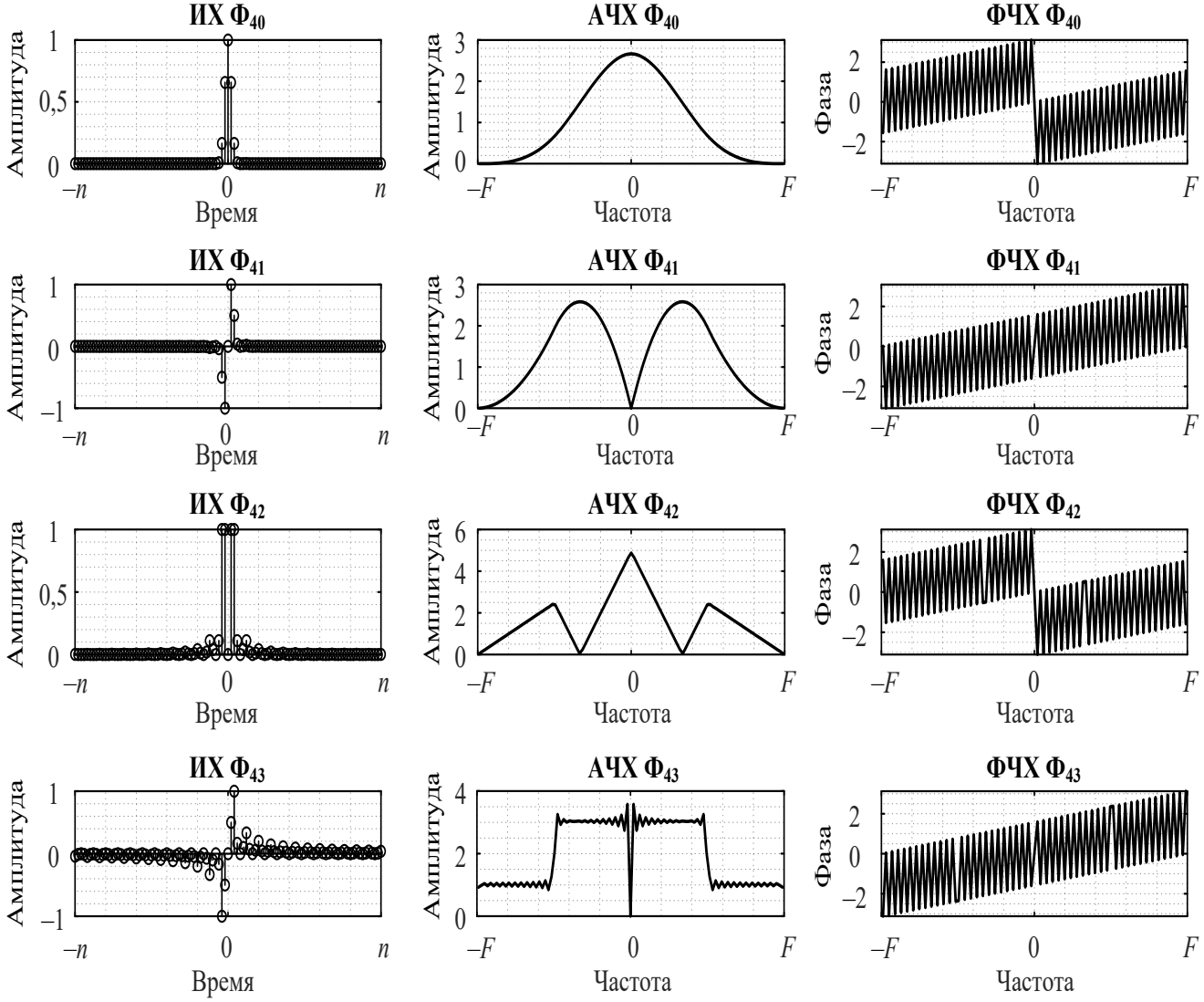


Рисунок 1.11 – АЧХ и ФЧХ синтезирующих фильтров представления Хургина – Яковлева при четырехканальной реализации

Также проведена модификация синтезирующих фильтров для представления Хургина – Яковлева при $K = 4$. АЧХ $A(f)$ и ФЧХ $\phi(f)$ синтезирующих фильтров в четырехканальной системе $(K = 4)$ имеют следующий вид [106]:

$$A_{40}(f) = 2{,}65 + 0{,}71(f / F) - 21{,}79(f / F)^2 + 31{,}54(f / F)^3 - 13{,}13(f / F)^4,$$
$$\phi_{40}(f) = 0. \qquad (1.34)$$

$$A_{41}(f) = -0{,}08 + 12{,}32(f / F) - 25{,}68(f / F)^2 + 12{,}22(f / F)^3 - 1{,}34(f / F)^4,$$
$$\phi_{41}(f) = \begin{cases} -\pi / 2 \text{ при } f \le F, \\ \pi / 2 \text{ при } F < f \le 2F. \end{cases} \qquad (1.35)$$

$$A_{42}(f)=\begin{cases} -9,72(f/F) \text{ при } f \le 0,33F, \\ 9,33(f/F)-3,05 \text{ при } 0,33F < f \le 0,5F, \\ -3,24(f/F)+3,24 \text{ при } 0,5F < f \le F. \end{cases}$$

$$\phi_{42}(f)=\begin{cases} -\pi \text{ при } f \le F, \\ \pi \text{ при } F < f \le 2F. \end{cases} \qquad (1.36)$$

$$A_{43}(f)=\begin{cases} 3,09 \text{ при } f \le 0,5F, \\ 1,03 \text{ при } 0,5F < f \le F. \end{cases},$$

$$\phi_{43}(f)=\begin{cases} \pi/2 \text{ при } f \le F, \\ -\pi/2 \text{ при } F < f \le 2F. \end{cases} \qquad (1.37)$$

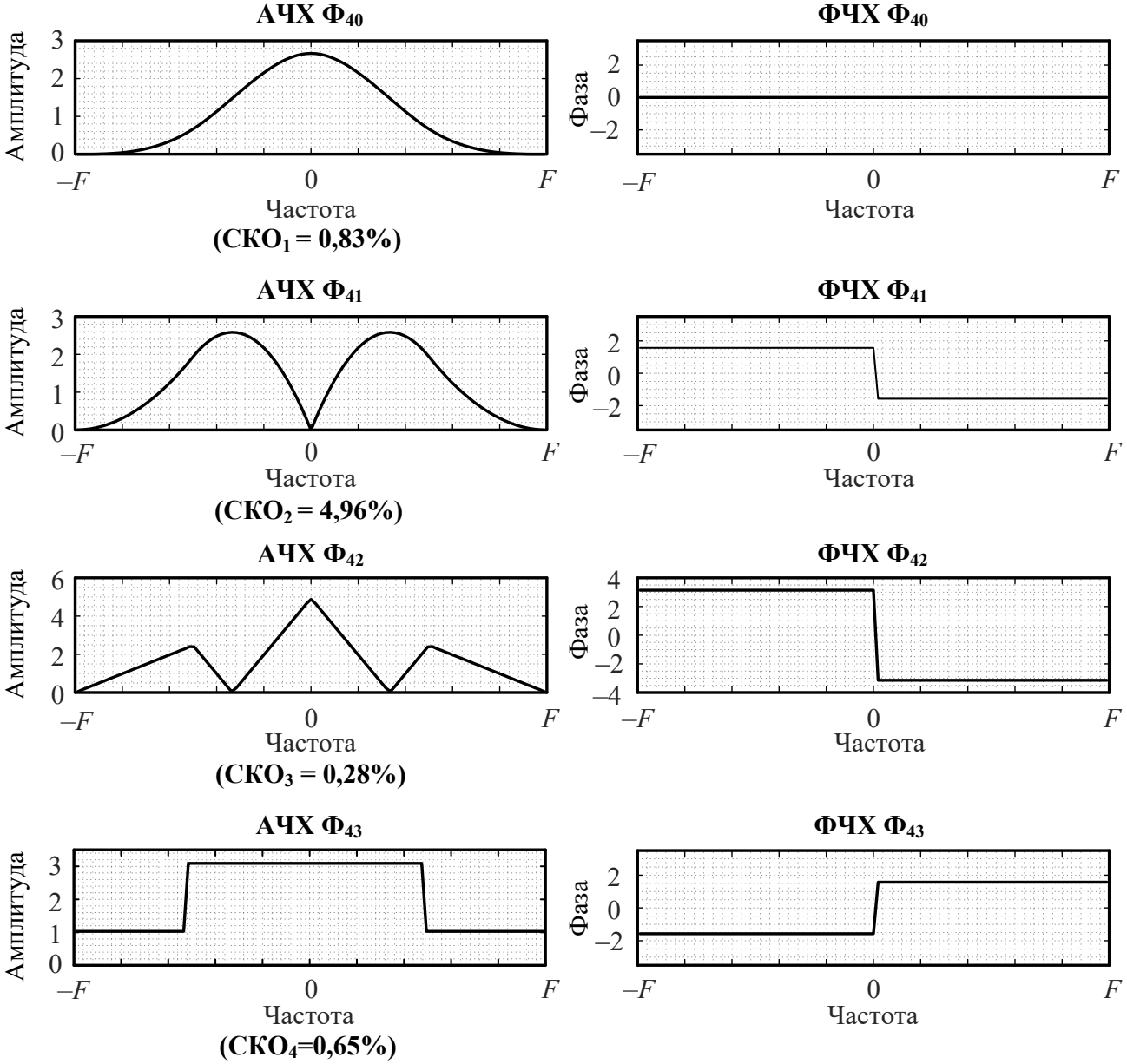


Рисунок 1.12 – АЧХ и ФЧХ синтезирующих фильтров при четырехканальной реализации после модификации

На основании синтезированных частотных характеристик построена структурная схема четырёхканальной системы разложения–восстановления, представленная на рисунке 1.13.

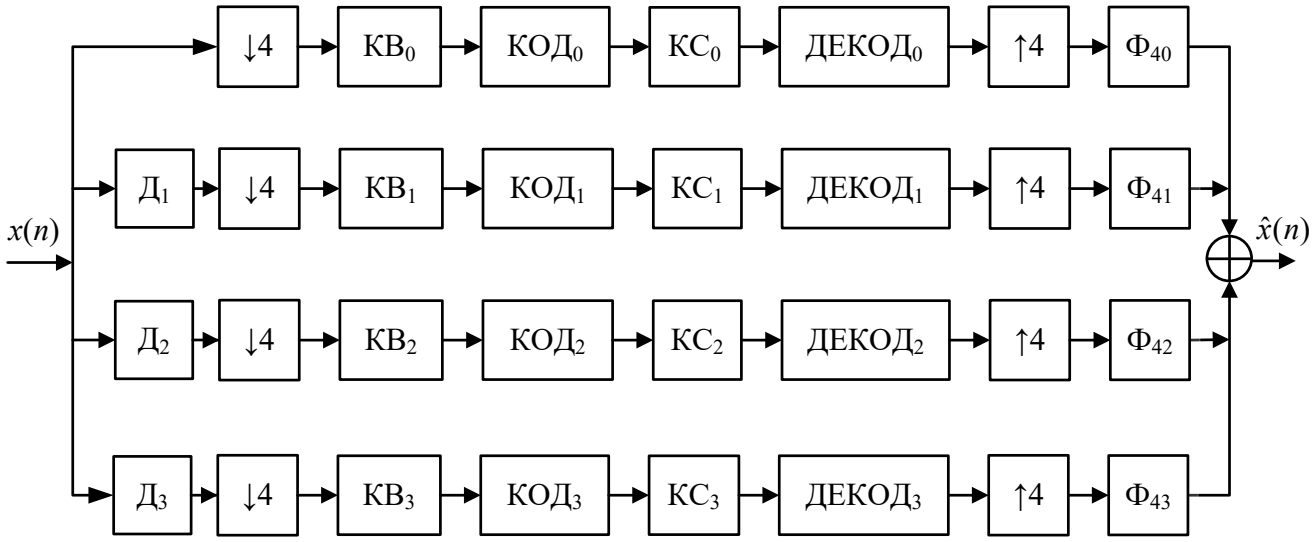


Рисунок 1.13 – Структурная схема системы обработки и передачи сигналов на основе представления Хургина – Яковлева при четырехканальной реализации

На передающей стороне входной сигнал $x(t)$, после дискретизации с частотой, определяемой теоремой В.А. Котельникова, поступает одновременно на четыре канала обработки: первый канал передаёт непосредственно отсчёты сигнала $x(n)$, а три остальных канала содержат дифференциаторы, вычисляющие производные первого, второго и третьего порядков соответственно $x'(n), x''(n), x'''(n)$. Все сформированные последовательности отсчётов сигнала и его производных затем подвергаются децимации с коэффициентом прореживания K=4 перед квантованием, кодированием и передачей в канал связи.

На стороне приёма каждая из четырёх последовательностей подаётся на соответствующий синтезирующий фильтр ($Ф_{40}$, $Ф_{41}$, $Ф_{42}$, $Ф_{43}$), после чего восстановленные частичные сигналы суммируются, образуя оценку $\hat{x}(t)$. Принципиальным отличием данной схемы от трёхканальной является наличие дополнительного блока вычисления третьей производной, что повышает требования к точности численного дифференцирования и непосредственно влияет на погрешность восстановления, исследуемую в разделе 1.5.

Таким образом, для K = 4 синтезированы АЧХ и ФЧХ всех четырёх

синтезирующих фильтров, построена структурная схема системы и определены погрешности аппроксимации. Полученные характеристики совместно с аналогичными результатами для $K = 2$ и $K = 3$ составляют основу для экспериментального исследования точности восстановления РС, проводимого в разделе 1.6.

1.5 Оценка точности алгоритмов вычисления производных сигнала для систем на основе представления Хургина – Яковлева

1.5.1 Требования к алгоритмам вычисления производных

Реализация представления Хургина – Яковлева в радиотехнических системах передачи РС требует решения важной практической задачи — обеспечения высокой точности вычисления производных дискретного сигнала. От выбора метода дифференцирования напрямую зависят как качество восстановленного сигнала, так и вычислительная сложность системы в целом. В данном разделе проводится сравнительный анализ существующих алгоритмов вычисления производных и формулируются рекомендации по их применению при реализации представления Хургина – Яковлева.

Дифференциатор в структурной схеме системы на основе представления Хургина – Яковлева обрабатывает отсчёты, взятые с частотой дискретизации $F_{\text{К}}$, и расположен перед дециматором. Его задача — формирование децимированных отсчётов производных, используемых на последующих этапах восстановления сигнала. Точность вычисления производных напрямую определяет качество восстановленного сигнала на выходе системы [107], поэтому выбор алгоритма дифференцирования является одной из ключевых задач практической реализации представления Хургина – Яковлева.

Специфика требований для РС:

Для РС с полосой частот 0…4 кГц частота дискретизации составляет $F_{\text{К}} = 8$ кГц. При использовании представления Хургина – Яковлева с различным количеством каналов возникают следующие требования к вычислению производных:

При $K = 2$: необходимо вычислять первую производную с шагом дискретизации $\Delta t = 1/F_{\text{К}} = 125$ мкс перед децимацией в 2 раза.

При $K = 3$: требуется вычисление первой и второй производных на частоте $F_К$ кГц перед децимацией в 3 раза, при этом точность вычисления второй производной критически важна.

При $K = 4$: необходимо вычисление трех первых производных на частоте $F_К$ перед децимацией в 4 раза, где точность третьей производной определяет применимость данного варианта реализации.

Формализация требований:

Обозначим через $\hat{x}'[n]$ отсчеты производной, вычисленные алгоритмом, а через $x'[n]$ – истинные значения производной. Тогда критерием точности алгоритма является нормированная среднеквадратическая ошибка (НСКО):

$$\sigma = \sqrt{\frac{\sum_{n=0}^{M-1}\left(\hat{x}'[n] - x'[n]\right)^2}{\sum_{n=0}^{M-1}\left(x'[n]\right)^2}}, \tag{1.38}$$

где M – количество отсчетов в анализируемом кадре сигнала.

Основные требования к алгоритмам:

1. Точность: НСКО σ должно быть минимальным при частоте дискретизации $F_Д = F_К$, без необходимости её повышения.

2. Вычислительная эффективность: количество операций на отсчет должно позволять обработку в реальном масштабе времени на цифровых сигнальных процессорах. Для кадров длиной M = 128...256 отсчетов (что соответствует 16…32 мс при $F_К = 8$ кГц) общее количество операций не должно превышать возможностей процессора при частоте обработки 50…100 Гц.

3. Устойчивость к шумам: алгоритм должен обеспечивать минимальную методическую погрешность аппроксимации характеристики $j\omega$ в рабочей полосе (0…4 кГц) и подавлять спектральные компоненты за её пределами для предотвращения усиления внеполосного шума и алиасинга.

4. Возможность параллельной реализации: для многоканальных систем представления Хургина – Яковлева желательна возможность одновременного

вычисления нескольких производных различных порядков, что снижает общие вычислительные затраты.

5. Совместимость с кадровой обработкой: поскольку речевые сигналы обрабатываются кадрами по 10…30 мс (квазистационарные участки), алгоритм должен корректно работать с кадрами конечной длины.

Далее будут рассмотрены основные методы вычисления производных дискретных сигналов и разработаны конкретные алгоритмы их реализации применительно к системам обработки на основе представления Хургина – Яковлева.

1.5.2 Методы вычисления производных дискретных сигналов и их применимость при реализации представления Хургина – Яковлева

Задача вычисления производных дискретных сигналов может быть решена различными методами, различающимися по точности и вычислительной сложности. Рассмотрим основные подходы применительно к обработке РС в системах на основе представления Хургина – Яковлева.

Метод конечных разностей.

Метод конечных разностей является наиболее простым подходом к численному дифференцированию. Его теоретическая основа базируется на классическом определении производной через предельный переход.

Математическая основа метода: Производная функции $x(t)$ в точке t определяется как:

$$x'(t_i) = \lim_{\Delta t \to 0} \frac{x(t+\Delta t) - x(t)}{\Delta t}. \tag{1.39}$$

При работе с дискретными отсчетами сигнала $x(t_i)$, взятыми с постоянным шагом Δt, непосредственное вычисление предела невозможно. Вместо этого метод конечных разностей использует аппроксимацию производной конечным отношением приращений при фиксированном значении Δt. Для вычисления производной в точке t_i могут быть использованы различные варианты разностных отношений.

Алгоритм правой разности:

$$x'(t_i) \approx \frac{x(t_{i+1}) - x(t_i)}{\Delta t}. \tag{1.40}$$

где $t_i = i\Delta t; i = 0,1,\ldots,K-1.$

Алгоритм левой разности:

$$x'(t_i) \approx \frac{x(t_i) - x(t_{i-1})}{\Delta t}. \tag{1.41}$$

Для анализа точности аппроксимации разложим функцию $x(t_i + 1)$ в ряд Тейлора в окрестности точки t_i:

$$x(t_{i+1}) = x(t_i) + x'(t_i)\Delta t + \frac{x''(t_i)}{2!}(\Delta t)^2 + \frac{x'''(t_i)}{3!}(\Delta t)^3 + \ldots \tag{1.42}$$

Подставляя (1.39) в выражение (1.37), получаем:

$$\frac{x(t_{i+1}) - x(t_i)}{\Delta t} = x'(t_i) + \frac{x''(t_i)}{2!}\Delta t + \frac{x'''(t_i)}{3!}(\Delta t)^2 + \ldots \tag{1.43}$$

Из выражения (1.40) следует, что главный член погрешности аппроксимации пропорционален Δt, что соответствует первому порядку точности $\mathrm{O}(\Delta t)$. Погрешность аппроксимации:

$$\varepsilon_{\text{кр}}(t_i) = \frac{x''(t_i)}{2}\Delta t + \mathrm{O}\left((\Delta t)^2\right). \tag{1.44}$$

Аналогичный анализ для левой разности дает ту же оценку порядка точности.

Вычислительная сложность алгоритмов: для вычисления производной в одной точке алгоритм требует: 1 операция вычитания; 1 операция деления (или умножения на предвычисленное значение $1/\Delta t$). Для обработки последовательности из *M* отсчетов общее количество операций составляет $2M$.

Метод полиномиальной интерполяции

Метод полиномиальной интерполяции представляет собой более совершенный подход, основанный на локальной аппроксимации сигнала гладкой функцией – полиномом – с последующим аналитическим дифференцированием этого полинома. Основная идея метода состоит в следующем: по набору из $n+1$ отсчетов сигнала строится интерполяционный полином $P_n(t)$ степени *n*, проходящий через все эти точки, после чего производная вычисляется как $x'(t_i) = P_n'(t_i)$. Для равноотстоящих узлов интерполяции, что имеет место при равномерной дискретизации сигнала, существуют эффективные рекуррентные

формулы, основанные на формулах Ньютона и Стирлинга, которые не требуют решения систем линейных уравнений.

Алгоритм центральной разности:

Простейшим случаем является использование трех точек $\{t_{i-1}, t, t_{i+1}\}$ для построения полинома второй степени. Применение формулы Лагранжа и дифференцирование приводит к классической формуле центральной разности:

$$x'(t_i) \approx \frac{x(t_{i+1}) - x(t_{i-1})}{2\Delta t}. \tag{1.45}$$

Разложим функции $x(t_{i+1})$ и $x(t_{i-1})$ в ряды Тейлора:

$$x(t_{i+1}) = x(t_i) + x'(t_i)\Delta t + \frac{x''(t_i)}{2!}(\Delta t)^2 + \frac{x'''(t_i)}{3!}(\Delta t)^3 + \dots \tag{1.46}$$

$$x(t_{i-1}) = x(t_i) - x'(t_i)\Delta t + \frac{x''(t_i)}{2!}(\Delta t)^2 - \frac{x'''(t_i)}{3!}(\Delta t)^3 + \dots$$

Вычитая второе уравнение из первого и деля на $2\Delta t$:

$$\frac{x(t_{i+1}) - x(t_{i-1})}{2\Delta t} = x'(t_i) + \frac{x'''(t_i)}{3!}(\Delta t)^2 + \mathrm{O}\left((\Delta t)^4\right) \tag{1.47}$$

Главный член погрешности:

$$\varepsilon_{\text{пр}}(t_i) = \frac{x'''(t_i)}{3!}(\Delta t)^2 + \mathrm{O}\left((\Delta t)^4\right), \tag{1.48}$$

что соответствует второму порядку точности метода $\mathrm{O}\left((\Delta t)^2\right)$. Это существенное улучшение по сравнению с односторонними разностями, так как погрешность уменьшается пропорционально квадрату шага дискретизации.

Для дальнейшего повышения точности можно использовать большее количество точек. Полином четвертой степени, построенный по пяти точкам $\{t_{i-2}, t_{i-1}, t, t_{i+1}, t_{i+2}\}$, дает следующую формулу для производной:

$$x'(t_i) \approx \frac{-x(t_{i+2}) + 8x(t_{i+1}) - 8x(t_{i-1}) + x(t_{i-2})}{12\Delta t}. \tag{1.49}$$

Анализ погрешности показывает, что главный член:

$$\varepsilon_{n4}(t_i) = \frac{x^{(5)}(t_i)}{30}(\Delta t)^4 + \mathrm{O}\left((\Delta t)^6\right), \tag{1.50}$$

что соответствует четвертому порядку точности $\mathrm{O}\left((\Delta t)^4\right)$.

Метод полиномиальной интерполяции может быть распространен на вычисление производных высших порядков путем многократного дифференцирования интерполяционного полинома.

Вторая производная (центральная формула):

$$x''(t_i) \approx \frac{x(t_{i+1}) - 2x(t_i) + x(t_{i-1})}{(\Delta t)^2}, \tag{1.51}$$

с погрешностью $\mathrm{O}\left((\Delta t)^2\right)$.

Третья производная (по четырем точкам):

$$x'''(t_i) \approx \frac{x(t_{i+2}) - 2x(t_{i+1}) + 2x(t_{i-1}) - x(t_{i-2})}{2(\Delta t)^3}. \tag{1.52}$$

Проведем анализ вычислительной сложности рассмотренных алгоритмов:

Для центральной разности (формула 1.45): необходимы 1 операция вычитания, 1 операция деления. Итого: 2 операции на отсчет.

Для полинома 4-й степени (формула 1.49): 2 операции умножения на коэффициенты, 3 операции сложения/вычитания, 1 операция деления. Итого: 6 операций на отсчет.

Для обработки последовательности из M отсчетов общее количество операций составляет $2M$ для центральной разности и $6M$ для полинома четвертой степени.

Метод в частотной области

Метод в частотной области представляет собой принципиально иной подход, основанный на преобразовании сигнала в частотную область, где операция дифференцирования выполняется алгебраически. Основным преимуществом метода является отсутствие методической погрешности аппроксимации – точность определяется только разрядностью вычислений.

Метод базируется на фундаментальном свойстве преобразования Фурье: операция дифференцирования во временной области эквивалентна умножению спектра сигнала на множитель $j\omega$ в частотной области. Для непрерывного сигнала $x(t)$ с преобразованием Фурье $X(\omega)$ справедливо соотношение:

$$F\left\{\frac{d^k x(t)}{dt^k}\right\} = (j\omega)^k X(\omega), \tag{1.53}$$

где k – порядок производной. Это свойство позволяет заменить операцию дифференцирования во временной области на простое умножение в частотной области.

Для дискретного сигнала $\{x(t_0), x(t_1), \ldots, x(t_{k-1})\}$, где $t_i = i\Delta t$, процедура вычисления производной состоит из трех этапов.

Этап 1. Вычисление спектра сигнала.

Применяется прямое БПФ к последовательности отсчетов:

$$S_c[n] = \text{БПФ}\{x(t_i)\},\ n = 0,1,\ldots,M-1, \tag{1.54}$$

где M – размер кадра.

Каждая спектральная компонента $S_c[n]$ соответствует физической частоте:

$$\omega_n = \frac{2\pi n}{M\Delta t}. \tag{1.55}$$

При этом индекс n от 0 до $M/2$ соответствует положительным частотам от 0 до $F_К/2$, а индексы от $M/2+1$ до $M-1$ представляют отрицательные частоты (для действительных сигналов).

Этап 2. Формирование спектра производной.

Спектр k-й производной формируется умножением спектра сигнала на соответствующую степень $j\omega_n$:

$$S_{\text{пр}}^{(k)}[n] = (j\omega_n)^k S_c[n]. \tag{1.56}$$

Для действительных РС необходимо учитывать следующие особенности:

- Для $n = 0$ (постоянная составляющая): $\omega_0 = 0$ следовательно $S_{\text{пр}}^{(k)}[0] = 0$ для любого $k \geq 1$.

- Для $n = M/2$ (частота Найквиста при четном M): $S_{\text{пр}}^{(k)}[M/2] = 0$.

- Для $n > M/2$ используется представление отрицательных частот: $\omega_n = 2\pi(n-M)/(M\Delta t)$.

Детализация для производных различных порядков:

Для первой производной $(k=1)$: $S_{\text{пр}}^{(1)}[n] = j\omega_n S_c[n]$.

Для второй производной $(k=2)$: $S_{\text{пр}}^{(2)}[n] = (j\omega_n)^2 S_c[n] = -\omega_n^2 S_c[n]$.

Для третьей производной $(k=3)$: $S_{\text{пр}}^{(3)}[n] = (j\omega_n)^3 S_c[n] = -j\omega_n^3 S_c[n]$.

Этап 3. Вычисление отсчетов производной через обратное БПФ:

Применяется обратное БПФ к спектру производной:

$$x^{(k)}(t_i) = \text{ОБПФ}\left\{S_{\text{пр}}^{(k)}[n]\right\}. \tag{1.57}$$

Вычислительная сложность алгоритма:

Для вычисления одной производной алгоритм требует: Прямое БПФ: $(M/2)\log_2 M$ комплексных умножений; Умножение на $j\omega$: M комплексных операций; Обратное БПФ: $(M/2)\log_2 M$ комплексных умножений. Общее количество комплексных умножений: $M + M\log_2 M$. Одно комплексное умножение эквивалентно 4 вещественным операциям (умножениям и сложениям), следовательно: $4M(\log_2 M + 1)$.

1.5.3 Экспериментальные исследования точности алгоритмов вычисления производных

Для объективной оценки точности рассмотренных алгоритмов и определения их применимости в системах на основе представления Хургина – Яковлева проведены экспериментальные исследования на синтетических сигналах с известной аналитической производной и на реальных речевых сигналах. Основной целью исследований является выбор оптимального алгоритма для различных вариантов реализации систем $(K=2,3,4)$, определение границ их применимости и получение количественных характеристик точности для практического проектирования.

Методика экспериментальных исследований:

В качестве тестовых сигналов использовались суммы гармонических компонент с различными частотами и амплитудами, что позволяет моделировать спектральные свойства РС и обеспечивает возможность точного вычисления эталонных значений производных. Исходный тестовый сигнал представлялся в виде:

$$x(t) = \sum_{n=0}^{V} a_n \sin\left(\omega_n t + \varphi_n\right), \tag{1.58}$$

где V – количество гармонических компонент, варьировалось от 10 до 50 для различных тестов; a_n – амплитуды, распределенные по равномерному закону в диапазоне $[0{,}5;1{,}5]$; $\omega_n = 2\pi f_n$, где частоты f_n равномерно распределены в диапазоне 100…4000 Гц, охватывающем полосу телефонного качества речи; φ_n – начальные фазы, распределенные равномерно в интервале $[0;2\pi]$. Для такого сигнала производная *k*-го порядка вычисляется аналитически:

$$x^{(k)}(t) = \sum_{n=0}^{V} a_n \omega_n^k \sin\left(\omega_n t + \varphi_n + k\frac{\pi}{2}\right), \tag{1.59}$$

что позволяет точно определить эталонные значения для сравнения с результатами работы алгоритмов.

Параметры экспериментов были выбраны следующим образом: количество независимых реализаций составило 500 для обеспечения статистической достоверности результатов; длина каждой реализации $M = 256$ отсчетов, что соответствует 32 мс при частоте дискретизации 8 кГц и согласуется с типичной длительностью квазистационарных участков речи; частота дискретизации $F_{\text{Д}}$ варьировалась от $0{,}7F_{\text{К}}$ до $1{,}5F_{\text{К}}$ для исследования зависимости точности от соотношения частот. Точность алгоритмов оценивалась путём вычисления НСКО по формуле (1.38) для каждой реализации с последующим усреднением результатов.

Сравнительный анализ алгоритмов при различных частотах дискретизации:

Экспериментально измеренные значения НСКО σ (в процентах) для различных алгоритмов вычисления первой производной при различных нормированных частотах дискретизации $F_{\text{Н}} = F_{\text{Д}}/F_{\text{К}}$ представлены на рисунке 1.14.

Рисунок 1.14 наглядно демонстрирует характер зависимости точности от частоты дискретизации.

Кривая 1 соответствует алгоритму конечных разностей и показывает наихудшую точность во всем диапазоне частот: $\sigma = 3{,}2\%$ при $F_{\text{Н}} = 1$, что

обусловлено принципиальной ограниченностью первым порядком точности $O(\Delta t)$. Даже при увеличении частоты до $F_{Н} = 1{,}5$ СКО снижается незначительно до 2,6%, что подтверждает фундаментальные ограничения простейших разностных схем.

Кривая 2 представляет метод полиномиальной интерполяции, использующий центральные разности по пяти точкам с полиномом четвертой степени. Благодаря высокому порядку аппроксимации $O\left((\Delta t)^4\right)$ данный метод показывает заметно лучшие результаты: $\sigma = 2{,}2\%$ при $F_{Н} = 1$, что на 31% лучше конечно-разностного метода. При повышении частоты дискретизации наблюдается существенное улучшение точности до 1,0% при $F_{Н} = 1{,}5$, что подтверждает эффективность высокопорядковых схем при передискретизации.

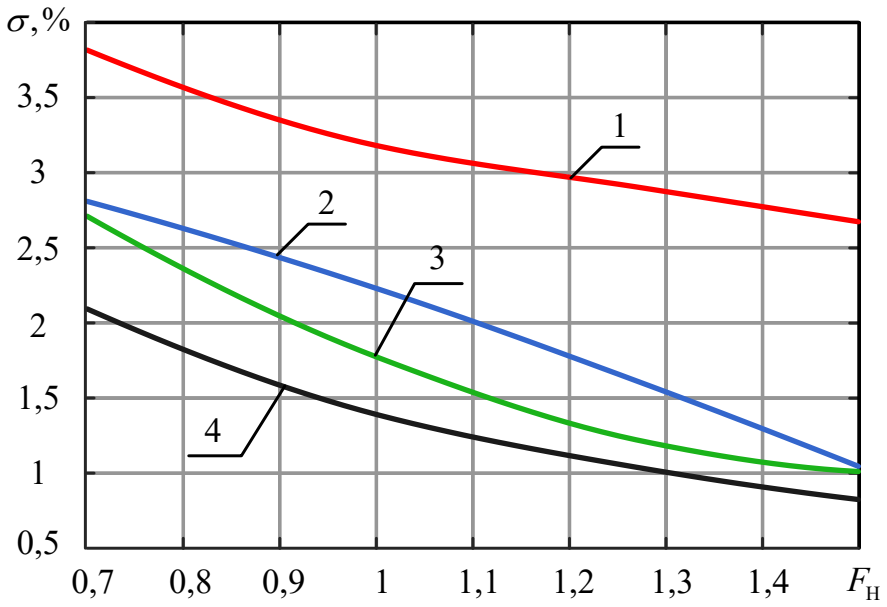


Рисунок 1.14 – Зависимость НСКО от нормированной частоты дискретизации для различных алгоритмов вычисления производных

Кривая 3 соответствует спектральному методу на основе БПФ. При $F_{Н} = 1$ метод обеспечивает $\sigma = 1{,}8\%$, что на 18% лучше полиномиального метода и на 44% лучше метода конечных разностей. Данное преимущество обусловлено отсутствием методической погрешности аппроксимации дифференциального оператора в частотной области.

Кривая 4 показывает теоретический предел точности при аналитическом вычислении производной тестового сигнала. При $F_{Н} = 1$ НСКО составляет 1,4%,

что определяется исключительно ошибками дискретизации и конечностью длины кадра. Разность между спектральным методом и аналитическим пределом составляет всего 0,4%, что свидетельствует о высокой близости спектрального метода к теоретическому оптимуму.

Исследование точности производных высших порядков

Для систем при $K \geq 3$ критически важна точность вычисления производных высших порядков. На рисунке 1.15 представлены зависимости НСКО σ от частоты дискретизации $F_Д$ для спектрального алгоритма: кривая 1 – первая производная $x'(t)$, кривая 2 – вторая производная $x''(t)$, кривая 3 – третья производная $x'''(t)$. Анализ результатов показывает, что при частоте дискретизации $F_Д = 8$ кГц достигнуты следующие значения НСКО: для первой производной $\sigma_1 \approx 0{,}01$, для второй производной $\sigma_2 \approx 0{,}01$, для третьей производной $\sigma_3 \approx 0{,}04$.

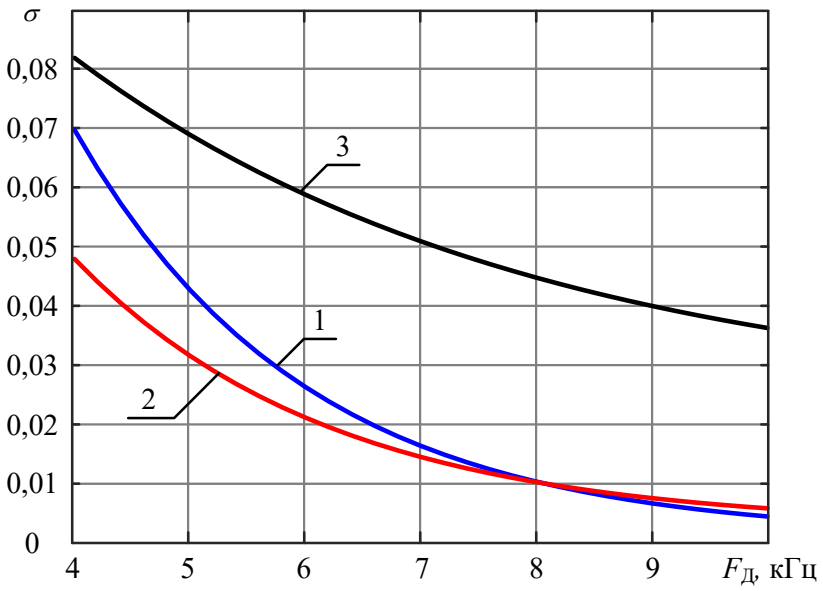


Рисунок 1.15 – Точность вычисления производных высших порядков спектральных алгоритмом

Высокая точность первой и второй производных обеспечивает применимость спектрального алгоритма в системах при $K = 2$ и $K = 3$. Несмотря на теоретическое усиление высокочастотных компонент пропорционально ω^2 для второй производной, практическая точность остается достаточной благодаря оптимальному проектированию фильтров, минимизирующему ошибку аппроксимации в рабочем частотном диапазоне.

Повышенное значение НСКО третьей производной объясняется усилением высокочастотных компонент пропорционально ω^3, что приводит к большей чувствительности к эффектам конечности сегмента и ошибкам дискретизации исходного сигнала. Тем не менее, достигнутая точность позволяет применение систем при $K = 4$ в низкоскоростных кодеках речи, где допускаются большие искажения при приоритете высокой степени сжатия.

Таким образом, результаты экспериментальных исследований подтверждают, что спектральный метод на основе БПФ обеспечивает наилучшие показатели точности среди рассмотренных алгоритмов при рабочей частоте дискретизации $F_Д = F_К = 8$ кГц, что делает его предпочтительным выбором для реализации систем на основе представления Хургина – Яковлева при всех рассмотренных значениях числа каналов $K = 2, 3, 4$.

1.6 Оценка точности восстановления сигналов на основе представления Хургина – Яковлева

1.6.1 Анализ составляющих погрешности восстановления сигнала

Анализ точности восстановления сигналов является одной из ключевых задач при практическом применении представления Хургина – Яковлева. Как было показано в предыдущих разделах, рассматриваемая реализация обеспечивает возможность параллельной обработки децимированных отсчетов сигнала и его производных при различном количестве каналов K. Однако практическое воплощение метода связано с возникновением погрешностей, непосредственно влияющих на качество восстановленного сигнала. При использовании многоэтапных структур разложения с числом этапов P накопление погрешностей на каждом из уровней дополнительно влияет на итоговое качество.

В данном разделе проводится комплексный анализ точности восстановления РС в идеальных условиях передачи, исключающих влияние шумов в канале связи, а также искажений, вносимых процедурами квантования и кодирования. Это позволяет выделить и оценить погрешности, обусловленные исключительно особенностями самого алгоритма обработки. Такой подход важен

для понимания фундаментальных ограничений метода и определения области его эффективного применения.

Как было показано в работах Я.И. Хургина и В.П. Яковлева [42], представление Хургина – Яковлева в общем виде (1.6) не обеспечивает точного восстановления сигнала с финитным спектром по причине неопределенности фазовых характеристик синтезирующих фильтров. При практической реализации представления Хургина – Яковлева погрешность восстановления определяется тремя основными компонентами:

$$\varepsilon_{\Sigma} = \varepsilon_{\phi} + \varepsilon_{der} + \varepsilon_{a}, \tag{1.60}$$

где $\varepsilon_{\phi}(t)$ – погрешность, обусловленная неопределенностью фазовой характеристики синтезирующих фильтров; $\varepsilon_{der}(t)$ – погрешность вычисления производной; $\varepsilon_{a}(t)$ – погрешность аппроксимации АЧХ синтезирующих фильтров.

Погрешность фазовой характеристики $\varepsilon_{\phi}(t)$ является фундаментальной для представления Хургина – Яковлева. В рамках разработанной модификации при K=2 эта погрешность полностью устраняется в диапазоне частот $\left(0;F\right)$ за счет использования синтезированных АЧХ и ФЧХ по формуле (1.11), (1.12). При рассмотрении полного частотного диапазона $\left[0;2F\right]$ погрешность концентрируется на частотах 0 и F. Для систем с $K \geq 3$ каналов погрешность минимизирована выбором ФЧХ в соответствии с выражениями (1.26)…(1.28) и (1.34)…(1.37).

Погрешность вычисления производной $\varepsilon_{der}(t)$ подробно проанализирована в разделе 1.5. Согласно результатам, при использовании алгоритма в частотной области НСКО производной составляет менее 2,7% в диапазоне нормированных частот $F_{\text{Н}} = 0,7 \ldots 1,5$. Для более высоких производных (второй и третьей) погрешность возрастает. Зависимость НСКО от частоты дискретизации, свидетельствует о том, что после частоты 8 кГц погрешность стабилизируется: для второй производной не более 0,01, для третьей – не более 0,04. При многоэтапном разложении $P > 1$ погрешности вычисления производной $\varepsilon_{der}(t)$ накапливаются аддитивно на каждом уровне.

Погрешность аппроксимации АЧХ $\varepsilon_a(t)$. При $K \geq 3$ невозможно получить аналитическое выражение для АЧХ всех синтезирующих фильтров, что требует применения аппроксимации полиномами. Для $K = 3$ АЧХ первого фильтра аппроксимируется полиномом четвертой степени (1.23), для $K = 4$ требуется аппроксимация АЧХ первого и второго фильтров выражениями (1.31), (1.32). Погрешность аппроксимации зависит от степени полинома и используемого критерия оптимизации. При применении критерия минимума среднеквадратического отклонения погрешность аппроксимации в диапазоне $[0; F]$ не превышает 7,69% для $K = 3$ и 4,96% для $K = 4$.

1.6.2 Методика оценки точности восстановления

Для комплексной оценки точности восстановления сигналов на основе представления Хургина – Яковлева была разработана методика экспериментальных исследований, включающая выбор тестового материала, набора объективных метрик качества и условий проведения эксперимента.

Тестовый материал. Экспериментальная оценка точности восстановления проводилась на представительном корпусе реальных РС. Для обеспечения репрезентативности и воспроизводимости результатов использовался набор из 100 записей от 10 дикторов различного пола в возрасте от 20 до 60 лет, в соответствии с требованиями ГОСТ Р 50840–95 [73]. Записи охватывали разнообразие фонетических контекстов: гласные, согласные, взрывные, фрикативные и аффрикативные звуки, а также слитную речь. Параметры оцифровки: частота дискретизации 8 кГц, разрядность 16 бит. Обработка вёлась по непересекающимся кадрам длиной $M = 256$ отсчётов (32 мс).

Объективные метрики качества

Для оценки точности восстановления используются три метрики из комплекса, определённого в разделе 1.2.2: SNR (формула (1.1)), MSE (формула (1.3)) и ViSQOL. Пороговые значения и порядок вычисления каждой метрики приведены в разделе 1.2.2.

Параметры эксперимента. Для каждой из 9 комбинаций параметров

$\left(K \in \{2;3;4\}, P \in \{1;2;3\}\right)$ выполнялось 500 независимых испытаний на случайно выбранных кадрах из корпуса. Обработка осуществлялась на платформе MATLAB R2023b. Передача сигналов производилась по идеальному каналу без квантования, шумов и нелинейных искажений, что позволяет изолировать погрешности, вносимые исключительно алгоритмом обработки.

Статистическая обработка. Для каждой конфигурации (*K*, *P*) по результатам 500 испытаний вычислялись среднее значение и доверительный интервал с уровнем значимости $\alpha = 0{,}01$ (доверительная вероятность 99%) в соответствии с нормальным законом распределения. Значения, приведённые в таблице 1.1, представляют собой средние по корпусу при данном уровне значимости. Для оценки устойчивости результатов использовался критерий Колмогорова–Смирнова для проверки нормальности распределения погрешностей.

1.6.3 Экспериментальные исследования точности восстановления

В таблице 1.1 представлены результаты экспериментальной оценки точности восстановления РС для всех исследуемых конфигураций.

Таблица 1.1 – Оценки точности восстановления РС при реализации дискретизации и обработки на основе представления Хургина – Яковлева с применением многоэтапного разложения

Числа каналов	Числа этапов	Оценка качества		
		MSE	SNR (дБ)	ViSQOL (Балл)
K = 2	*P* = 1	$7{,}6.10^{-12}$	85,38	4,73
	P = 2	$9{,}9.10^{-9}$	55,93	4,72
	P = 3	$3{,}7.10^{-8}$	48,88	4,69
K = 3	*P* = 1	$6{,}3.10^{-6}$	26,02	4,71
	P = 2	$8{,}7.10^{-6}$	24,47	4,65
	P = 3	$9{,}3.10^{-6}$	24,04	4,50
K = 4	*P* = 1	$3{,}6.10^{-5}$	17,92	4,69
	P = 2	$4{,}5.10^{-5}$	15,66	4,64
	P = 3	$5{,}9.10^{-5}$	14,03	4,30

Двухканальная система $K=2$ при одноэтапном разложении $(K=2,P=1)$ демонстрирует исключительно высокий показатель SNR = 85,38 дБ при MSE = $7{,}6\cdot10^{-12}$ и ViSQOL = 4,73 балла. Столь высокое значение SNR объясняется совокупным действием нескольких факторов. Во-первых, при $K=2$ погрешность ε_ϕ полностью устранена за счёт применения аналитически синтезированных фильтров (1.8), (1.9), что принципиально отличает данную конфигурацию от всех остальных. Во-вторых, отсутствует погрешность аппроксимации АЧХ $(\varepsilon_a=0)$, поскольку для $K=2$ аналитические выражения для АЧХ обоих синтезирующих фильтров получены в замкнутой форме. В-третьих, суммарная погрешность определяется исключительно точностью вычисления первой производной $(\varepsilon_{der}\le 0{,}01)$. При увеличении числа этапов до $P=2$ и $P=3$ SNR снижается до 55,93 дБ и 48,88 дБ соответственно вследствие накопления ε_{der} по каждому уровню, однако ViSQOL практически не изменяется (4,72 балла и 4,69 балла), что свидетельствует о перцептивной устойчивости двухканальной системы.

Переход к трёхканальной системе $K=3$ сопровождается принципиальным изменением механизма формирования погрешности: возникает составляющая $\varepsilon_a\le 0{,}05\%$ и дополнительная составляющая ε_{der} второй производной (НСКО $\le$ 0,01). Это обусловливает значительное снижение SNR: при $P=1$ SNR = 26,02 дБ, что на 59,36 дБ меньше, чем для $K=2,P=1$. Тем не менее полученное значение превышает пороговое значение 13 дБ. Значение ViSQOL = 4,71 балла при $P=1$ остаётся на высоком уровне, что подтверждает применимость трёхканальной системы для практических задач. С ростом P наблюдается умеренная деградация: при $P=3$ SNR = 24,04 дБ (снижение на 1,98 дБ относительно $P=1$), ViSQOL = 4,50 балла. Важно отметить малую чувствительность SNR к изменению P при $K=3$ (диапазон 24…26 дБ), что свидетельствует о том, что при данном N доминирующей является погрешность аппроксимации ε_a, а не накопленная погрешность производной.

Четырёхканальная конфигурация $K=4$ при $P=1$ обеспечивает

SNR = 17,92 дБ, MSE = $3{,}6 \cdot 10^{-5}$ и ViSQOL = 4,69 балла. Снижение SNR обусловлено суммарным вкладом трёх составляющих: ε_a (до 4,96% при аппроксимации АЧХ фильтров Φ_{40}, Φ_{41}, Φ_{42} и Φ_{43}), ε_{der} второй производной (НСКО ≤ 0,01) и ε_{der} третьей производной (НСКО ≤ 0,04, усиленное множителем ω^3). Принципиальной особенностью данной конфигурации является высокое значение ViSQOL = 4,69 балла при $P = 1$. При $P = 3$ SNR снижается до 14,03 дБ, а ViSQOL = 4,30 балла, что остаётся выше порога приемлемости для телефонных систем.

Результаты исследований, приведенные в таблице 1.1, демонстрируют, что точность восстановления РС при использовании предложенного метода уменьшается с увеличением числа этапов и каналов обработки, а также при использовании отсчетов высших производных сигнала. Это видно из снижения оценок качества восстановленной речи при увеличении числа каналов *K* и порядка разложения *P*, что полностью согласуется с теоретическими положениями, описанными ранее. Помимо этого, экспериментальные данные подтверждают, что применение предложенного алгоритма в двухканальной структуре с двух- или трехэтапным разложением обеспечивает уровень SNR не ниже 45 дБ, что является достаточным для обработки РС и широкого спектра других типов данных. Показано, что использование трех- или четырехканальной структуры в сочетании с двух или трехэтапным разложением позволяет достичь показателя SNR в диапазоне 14 дБ и выше, что считается приемлемым для низкоскоростных систем передачи РС и ряда специализированных приложений. Применение систем передачи на основе представления Хургина – Яковлева при количестве каналов больше 4, т. е. $K > 4$ не целесообразно, так как SNR при этом будет еще ниже.

Для более детального анализа зависимостей на рисунках 1.16 и 1.17 представлены зависимости SNR и ViSQOL от числа этапов разложения *P* для всех рассмотренных конфигураций *K*. Сопоставление этих результатов позволяет сформулировать следующие ключевые выводы.

– Переход от *K* = 2 к *K* = 3 сопровождается значительным снижением SNR (в среднем на 59,4 дБ при *P* = 1), обусловленным принципиальным изменением

характера погрешности: при $K = 2$ ошибка аппроксимации отсутствует, тогда как при $K = 3$ и $K = 4$ доминирующей становится погрешность аппроксимации АЧХ фильтров. Переход от $K = 3$ к $K = 4$ сопровождается менее значительным снижением SNR — на 8,1 дБ при $P = 1$, поскольку оба случая ограничены одним и тем же механизмом погрешности.

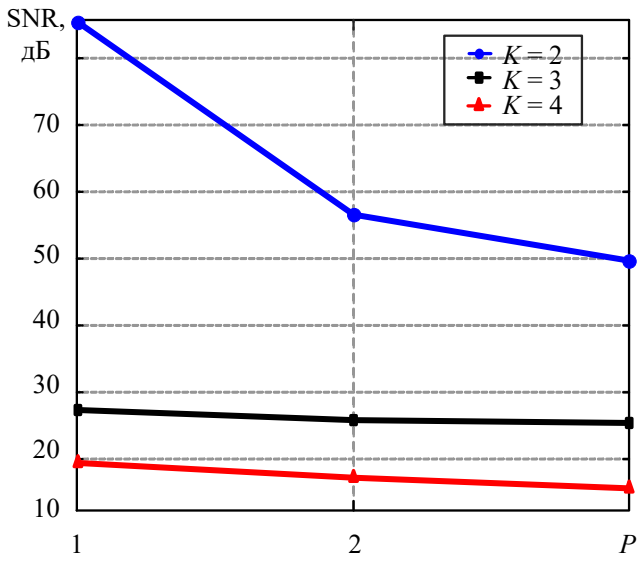


Рисунок 1.16 – Зависимость SNR от числа этапов многоэтапного разложения P при различном числе каналов K

– Для всех конфигураций K увеличение P приводит к монотонному снижению SNR. При $K = 2$ наблюдается наибольшее абсолютное падение — с 85,38 до 48,88 дБ при переходе от $P = 1$ к $P = 3$, тогда как для $K = 3$ и $K = 4$ снижение менее выражено и составляет 2…4 дБ.

На рисунке 1.17 показаны зависимость перцептивной оценки ViSQOL от числа этапов многоэтапного разложения P при различном числе каналов K.

Несмотря на значительный разброс SNR между конфигурациями $K = 2$ и $K = 4$ (более 67 дБ при $P = 1$), показатели ViSQOL изменяются незначительно: от 4,73 до 4,69 балла при $P = 1$, а при $P = 3$ — от 4,69 до 4,30 балла. Это свидетельствует о том, что погрешности восстановления сосредоточены преимущественно в спектральных областях с низкой чувствительностью слуха, и метод остаётся перцептивно адекватным даже при относительно низких энергетических метриках.

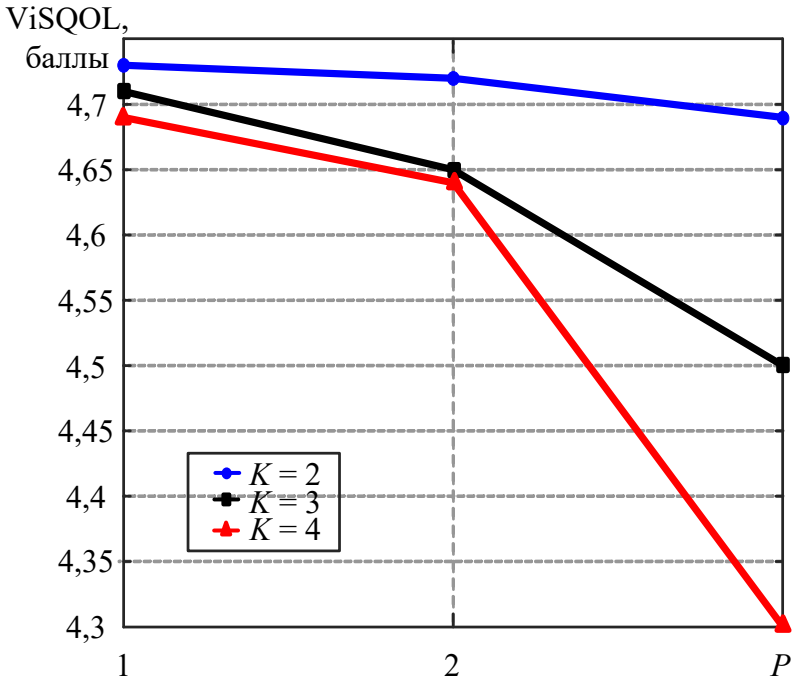


Рисунок 1.17 – Зависимость перцептивной оценки ViSQOL от числа этапов многоэтапного разложения P при различном числе каналов K

На основании проведённого анализа можно выделить режим работы системы. Для применений с высокими требованиями к качеству рекомендуется конфигурация $K = 2$, $P = 1$, обеспечивающая двукратную децимацию при качестве, близком к эталонному.

1.7 Выводы

В первой главе проведён комплексный анализ теоретических основ первичного кодирования РС на базе представления Хургина – Яковлева, синтезированы фильтры систем разложения–восстановления, исследованы методы вычисления производных и дана экспериментальная оценка точности восстановления. По результатам выполненных исследований сформулированы следующие выводы.

1. Анализ существующих методов дискретизации выявил принципиальные ограничения классического подхода В.А. Котельникова применительно к речевым сигналам: необходимость строгой полосовой ограниченности, высокая частота дискретизации, избыточность при передаче коррелированных участков речи. Спектрально-статистические характеристики речи подтверждают применимость

представления Хургина – Яковлева и обосновывают целесообразность использования систем с числом каналов $K = 2…4$.

2. На основе представления Хургина – Яковлева обоснован подход к первичному кодированию РС, при котором использование отсчётов сигнала и его производных позволяет сократить число передаваемых отсчётов в K раз по сравнению с классической дискретизацией, что теоретически обосновывает возможность снижения битрейта системы кодирования.

3. Синтезированы частотные характеристики фильтров систем разложения–восстановления представления Хургина – Яковлева для $K = 2, 3, 4$ каналов и числа этапов разложения $P = 1, 2, 3$. Для двухканальной системы ($K = 2$) доказано условие точного восстановления сигнала. Для систем $K = 3$ и $K = 4$ получены приближённые характеристики методом кусочно-линейной аппроксимации.

4. Проведён сравнительный анализ трёх методов вычисления производных в дискретной области: метода конечных разностей, полиномиальной интерполяции и метода на основе частотной области. Установлено, что метод на основе частотной области обеспечивает наилучшую точность вычисления производных при НСКО $\sigma \le 0{,}01$ и принят в качестве базового для реализации представления Хургина – Яковлева.

5. Экспериментальная оценка точности восстановления РС в условиях идеального канала (без квантования и шума) подтвердила алгоритмическую эффективность предложенного подхода. Для двухканальной системы $K = 2$ при $P = 1$ достигнуто SNR = 85,38 дБ и ViSQOL = 4,73 балла. Двухканальная структура при $P = 1…2$ обеспечивает SNR не ниже 45 дБ, что достаточно для широкого спектра задач обработки речи. Трёх- и четырёхканальные структуры обеспечивают SNR в диапазоне 14 дБ и выше при сохранении высоких перцептивных показателей (ViSQOL ≥ 4,69 балла), что приемлемо для низкоскоростных систем. Применение систем с $K > 4$ нецелесообразно ввиду дальнейшего падения SNR без существенного выигрыша в степени компрессии.

ГЛАВА 2 РЕАЛИЗАЦИЯ АЛГОРИТМОВ ПЕРВИЧНОГО КОДИРОВАНИЯ РЕЧЕВЫХ СИГНАЛОВ НА ОСНОВЕ ПРЕДСТАВЛЕНИЯ ХУРГИНА – ЯКОВЛЕВА

2.1 Вводные замечания

Развитие радиотехнических систем передачи РС неразрывно связано с поиском методов, обеспечивающих приемлемое качество восстановленной речи при минимальном использовании ресурсов канала связи. Среди актуальных направлений особое место занимает задача первичного кодирования, от эффективности которого в значительной мере зависят как пропускная способность системы передачи, так и её устойчивость к помехам. Традиционные методы первичного кодирования, основанные на теореме В.А. Котельникова, обеспечивают надёжную работу лишь при строгих условиях ограниченности полосы сигнала и высокой вычислительной нагрузке. В условиях ограниченных технических ресурсов, характерных для встраиваемых радиотехнических систем, возникает фундаментальное противоречие между требованиями к качеству речи и вычислительными возможностями аппаратных средств [15–20].

Представление Хургина – Яковлева [11] предоставляет альтернативный математический аппарат, в котором дискретизация сигнала осуществляется с привлечением отсчётов производных различных порядков. Это позволяет снизить частоту дискретизации в K раз по сравнению с классическим подходом при сохранении условий точного восстановления речевого сигнала [12, 13]. Практическая значимость данного подхода для задач первичного кодирования речи определяется двумя принципиальными факторами. Во-первых, для передачи отсчетов производной возможно использовать меньшую разрядность отсчетов, что уменьшает требуемую скорость передачи и расширяет возможности применения стандартных кодеков — MMBE, G.723.1, G.726 — в условиях ограниченной пропускной способности канала. Во-вторых, фазовые характеристики синтезирующих фильтров, обусловленные использованием производных, создают предпосылки для снижения влияния канального шума и шумов квантования на качество восстановленной речи [106, 107].

Вместе с тем, полноценная практическая реализация представления Хургина – Яковлева в системах первичного кодирования требует решения задач, которые в известной литературе не получили широкого распространения. С точки зрения вычислительной эффективности, потенциал снижения числа операций реализуется лишь при интерпретации алгоритма обработки в терминах теории многоскоростных банков фильтров с применением полифазного разложения [22, 22]. Без строгого определения эквивалентности представления Хургина – Яковлева и многоканального банка фильтров анализа-синтеза, а также без аналитических оценок числа операций при одноэтапном и многоэтапном разложении для различных конфигураций вопрос о реализуемости системы в реальном времени остаётся открытым. С точки зрения помехоустойчивости, фазовый сдвиг частотных характеристик синтезирующих фильтров теоретически должен обеспечивать некогерентное сложение шумовых составляющих на выходе сумматора и, следовательно, выигрыш в отношении сигнал-шум [107]. Однако аналитическая оценка величины этого выигрыша в зависимости от параметров K и P, а также его подтверждение в условиях битовых ошибок реального канала связи применительно к конкретным первичным кодекам в литературе не представлены в полном объёме. Таким образом, несмотря на имеющиеся теоретические предпосылки, практическая эффективность представления Хургина – Яковлева в качестве основы систем первичного кодирования РС не обоснована в достаточной мере.

Предметом настоящей главы является разработка и исследование алгоритмов первичного кодирования РС на основе представления Хургина – Яковлева с учётом вычислительных ограничений и реальных условий канала связи, включая воздействие аддитивного шума и битовых ошибок.

В настоящей главе решаются следующие задачи:

1. Установление эквивалентности представления Хургина – Яковлева многоканальному банку фильтров анализа-синтеза, построение полифазной структуры и получение аналитических выражений для числа вычислительных операций при одноэтапном и многоэтапном разложении в зависимости от числа каналов $K = 2, 3, 4$ и числа этапов $P = 1, 2, 3$.

2. Аналитическая оценка помехоустойчивости алгоритма при воздействии аддитивного белого гауссовского шума и определение выигрыша по сравнению с системой на основе теоремы В.А. Котельникова для всех рассматриваемых конфигураций, включая многоэтапное разложение.

3. Исследование устойчивости алгоритмов первичного кодирования к ошибкам в канале связи при вероятности битовых ошибок $P_{ош}$ = 0,1…5% применительно к первичным кодекам MMBE (1,2 кбит/с), G.723.1 (6,3 кбит/с) и G.726 (24 кбит/с).

4. Сравнительная экспериментальная оценка эффективности предложенных алгоритмов первичного кодирования по метрикам SNR и ViSQOL в сопоставлении с системой на основе теоремы В.А. Котельникова при тех же условиях кодирования и передачи.

2.2 Реализация представления Хургина – Яковлева в виде банка фильтров и оценка вычислительной сложности

2.2.1 Эквивалентное представление Хургина – Яковлева в форме многоканального банка фильтров анализа-синтеза

Реализация представления Хургина – Яковлева в цифровых радиотехнических системах передачи РС предполагает прежде всего выбор вычислительной архитектуры, при которой потенциальные преимущества метода — снижение частоты дискретизации и параллельная обработка каналов — могут быть реализованы в условиях ограниченных ресурсов встраиваемых систем. Как будет показано ниже, такая архитектура определяется теорией многоскоростных банков фильтров: именно интерпретация представления Хургина – Яковлева в терминах анализирующих и синтезирующих фильтров открывает путь к применению полифазного разложения и, как следствие, к кардинальному снижению числа вычислительных операций.

Общая модель K-канального банка фильтров. В *K*-канальном банке фильтров с максимальной децимацией входной сигнал $x(n)$ с помощью K анализирующих фильтров $H_k(z), k = 0,1,\ldots,K-1,$ разбивается на K канальных сигналов, каждый из которых подвергается K-кратной децимации. В блоке синтеза частота дискретизации восстанавливается в K раз интерполяторами с последующей

фильтрацией синтезирующими фильтрами $F_k(z)$ и суммированием.

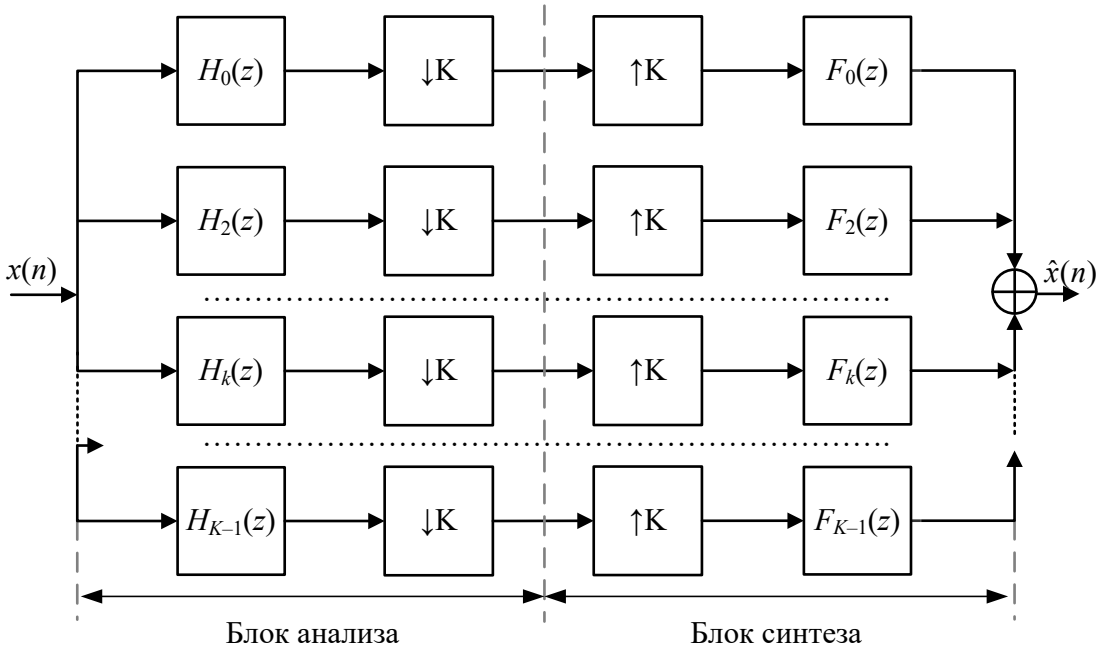


Рисунок 2.1 – Структура *K*-канального банка фильтров анализа-синтеза с максимальной децимацией

Z-преобразование восстановленного сигнала на выходе системы описывается выражением [22]:

$$\hat{X}(z)=\frac{1}{K}\sum_{n=0}^{K-1}X\left(zW^n\right)\sum_{k=0}^{K-1}H_k\left(zW^n\right)F_k(z),\ W=e^{-j2\pi/K}. \tag{2.1}$$

Слагаемые при $n>0$ в первой сумме описывают эффекты алиасинга (наложения спектров), неизбежно возникающие при наличии дециматоров и интерполяторов. Полная передаточная функция системы от входа к выходу:

$$T(z)=\frac{\hat{X}(z)}{X(z)}=\frac{1}{K}\sum_{k=0}^{K-1}H_k(z)F_k(z). \tag{2.2}$$

Если $T(z)=cz^{-k_0}$ для некоторого целого $k_0\geq 0$ и вещественной константы *c*, банк фильтров является системой с полным восстановлением: выходной сигнал является задержанной версией входного. В противном случае функция *T*(*z*) описывает амплитудные и фазовые искажения, претерпеваемые сигналом в структуре, и называется функцией искажений. Необходимым и достаточным условием устранения алиасинга является выполнение для всех $n\neq 0$ следующего условия:

$$\sum_{k=0}^{K-1}H_k\left(zW^n\right)F_k(z)=0,\ n=1,2,\ldots,K-1. \tag{2.3}$$

Условия (2.2), (2.3) допускают единое матричное представление $\mathbf{H}(z)f(z)=K\left[cT(z)\ 0\ \ldots\ 0\right]^T$, где $\mathbf{H}(z)$ — матрица составляющих наложения размером $K\times K$ с элементами $H_k\left(zW^n\right)$, $f(z)=\left[F_0(z)\ \ \ldots\ \ F_{K-1}(z)\right]^T$ [26]. При обратимости матрицы $\mathbf{H}(z)$ из данного уравнения можно найти синтезирующие фильтры $F_k(z)$, при которых эффект наложения полностью устраняется.

Отображение представления Хургина – Яковлева на модель банка фильтров. Покажем, что система обработки РС на основе представления Хургина – Яковлева, структурные схемы которой рассмотрены в главе 1 (рисунки 1.1, 1.3, 1.6), является частным случаем *K*-канального банка фильтров с максимальной децимацией. В анализирующем блоке структурной схемы *k*-й канал реализует вычисление конечной разности порядка *k* с последующей *K*-кратной децимацией. В частотной области данной операции соответствует анализирующий фильтр [27]:

$$H_k(z)=\left(1-z^{-1}\right)^k,\ k=1,2,\ldots,K-1. \tag{2.4}$$

Данный выбор обусловлен принципиальным свойством конечных разностей: разность порядка *k* является численным аналогом *k*-й производной сигнала — именно отсчёты самого сигнала и отсчёты его производных образуют канальные сигналы в структурных схемах главы 1. Структурная схема эквивалентного банка фильтров представления Хургина – Яковлева приведена на рисунке 2.2.

Достижимость условия полного восстановления. Для набора фильтров (2.4) матрица наложения $\mathbf{H}(z)$ содержит элементы $H_k\left(zW^n\right)=\left(1-W^nz^{-1}\right)^k$. Поскольку значения $\left(1-W^nz^{-1}\right)$ при n = 0, 1, …, N−1 различны при любом $z\neq 0$, матрица $\mathbf{H}(z)$ является обратимой, и система уравнений (2.3) имеет решение. Следовательно, существуют синтезирующие фильтры $F_k(z)$, при которых условие полного восстановления дискретного сигнала выполняется при любом *K*. Явный вид этих фильтров и доказательство условия полного восстановления с помощью аппарата полифазного разложения приводятся в разделе 2.2.2.

Разграничение дискретного и непрерывного уровней задачи. Изложенный выше результат относится к дискретной постановке задачи. Реальная система на

основе представления Хургина – Яковлева решает принципиально более сложную задачу в непрерывной области: восстановление непрерывного сигнала $x(t)$ из выборок его производных, переданных по каналу связи. Синтезирующие фильтры $F_k(z)$ здесь проектируются с учётом непрерывных спектральных характеристик производных, что принципиально отличается от дискретного случая.

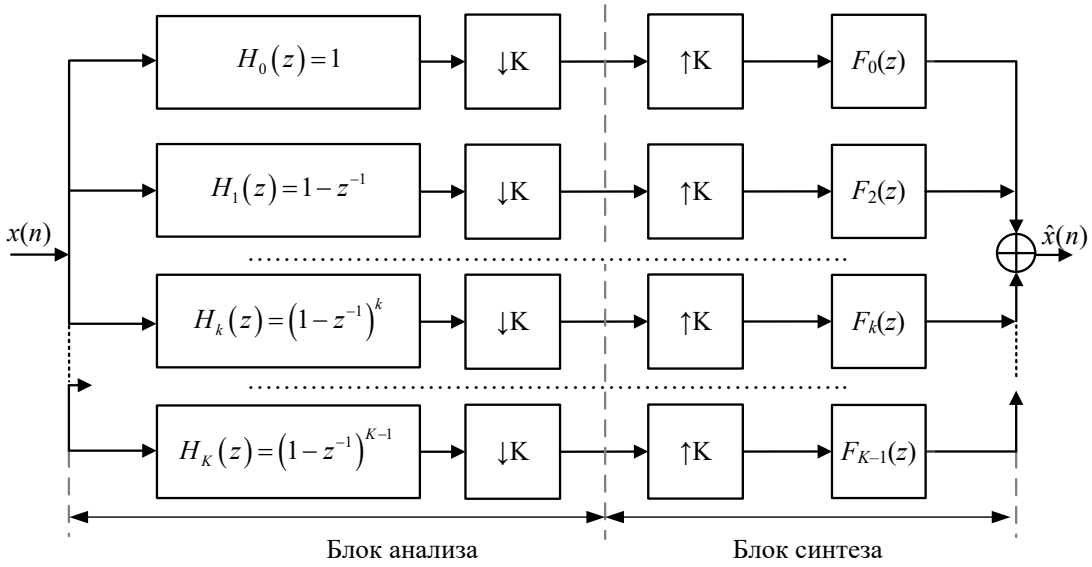


Рисунок 2.2 – Эквивалентное представление системы первичного кодирования РС на основе представления Хургина – Яковлева в форме K-канального банка фильтров анализа-синтеза

Для двухканальной системы (K = 2) система уравнений в непрерывной области (1.13), (1.14) имеет единственное решение с ненулевым определителем $\Delta = -j2\pi F^2 \neq 0,$ и синтезирующие характеристики (1.8), (1.9) обеспечивают точное восстановление непрерывного сигнала. Условие полного восстановления выполняется и в непрерывной области. Для трёх- и четырёхканальных систем (K = 3, 4) аналитическое решение для ФЧХ синтезирующих фильтров в непрерывной области в замкнутой форме недостижимо. Применение кусочно-линейной аппроксимации по критерию минимума СКО (раздел 1.4) приводит к тому, что условие полного восстановления в непрерывной области выполняется приближённо с погрешностями. Это ограничение обусловлено непрерывной постановкой задачи синтеза фильтров и не является нарушением условия полного восстановления в дискретной области, достижимого при любом K.

Таким образом, система на основе представления Хургина – Яковлева формально эквивалентна K-канальному банку фильтров анализа-синтеза с максимальной децимацией, анализирующие фильтры которого определяются выражением (2.4), а синтезирующие — характеристиками, полученными в разделе 1.4. Условие полного восстановления сигнала выполняется точно при $K = 2$ и приближённо при $K = 3, 4$. Полифазная реализация данной структуры, обеспечивающая заметное снижение вычислительных затрат, рассматривается в разделе 2.2.2.

2.2.2 Полифазное разложение банка фильтров представления Хургина – Яковлева

Полифазное представление анализирующих фильтров. Использование полифазного представления передаточных характеристик фильтров является ключевым инструментом построения эффективных структур банков фильтров [26]. Передаточную характеристику k-го анализирующего фильтра запишем в полифазной форме в следующем виде [26–28]:

$$H_k(z) = \sum_{l=0}^{K-1} E_{kl}\left(z^K\right) z^{-l}, \tag{2.5}$$

где $E_{kl}(z)$ – l-я полифазная компонента фильтра $H_k(z)$. Для анализирующих фильтров представления Хургина – Яковлева (2.4) раскрытие биномиальной формулы даёт:

$$E_{kl} = (-1)^l C_k^l, \tag{2.6}$$

где $C_k^l = \dfrac{k!}{l!(k-l)!}$ – биномиальный коэффициент. Таким образом, E_{kl} являются константами, не зависящими от z, и матрица полифазных компонент $\mathbf{E} = \left[E_{kl}\right]$ является постоянной матрицей биномиальных коэффициентов [9]:

$$\mathbf{E} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 1 & -2 & 1 & \dots & 0 \\ 1 & -3 & 3 & \dots & 0 \\ \vdots & & & \ddots & \vdots \end{bmatrix}. \tag{2.7}$$

Это принципиально отличает систему на основе представления Хургина – Яковлева от произвольных банков фильтров: обе полифазные матрицы — анализа и синтеза — являются матрицами констант, что исключает операции умножения на частоте F/K и оставляет только операции сложения.

Условие полного восстановления. Согласно [23], условие полного восстановления в полифазном представлении требует выполнения откуда $\mathbf{R} = \mathbf{E}^{-1}$. Биномиальная матрица (2.7) обладает свойством инволютивности [23]: $\mathbf{E}^{-1} = \mathbf{E}$, что доказывается прямым вычислением: каждый элемент произведения $\mathbf{E} \times \mathbf{E} = \mathbf{I}$ с использованием формулы свёртки биномиальных коэффициентов [23]. Следовательно, $\mathbf{R} = \mathbf{E}$ и справедливо следующее равенство:

$$\mathbf{P}(z) = \mathbf{E} \times \mathbf{E} = \mathbf{I}, \tag{2.8}$$

Условие полного восстановления дискретного сигнала выполняется тождественно при любом *K*. Данный факт является дискретным аналогом теоремы В.А. Котельникова о дискретизации производных непрерывных сигналов [9].

Построение полифазной структуры. Применяя тождества [24], матрицу **E** переносим за блоки дециматоров (↓K), а матрицу **R** = **E** — перед блоки интерполяторов (↑K). В результате вся обработка выполняется над последовательностями длиной F/K, вместо *F*. Полученная структура показана на рисунке 2.3. Аналитическая оценка вычислительного выигрыша, достигаемого данным разложением, приводится в разделе 2.2.3.

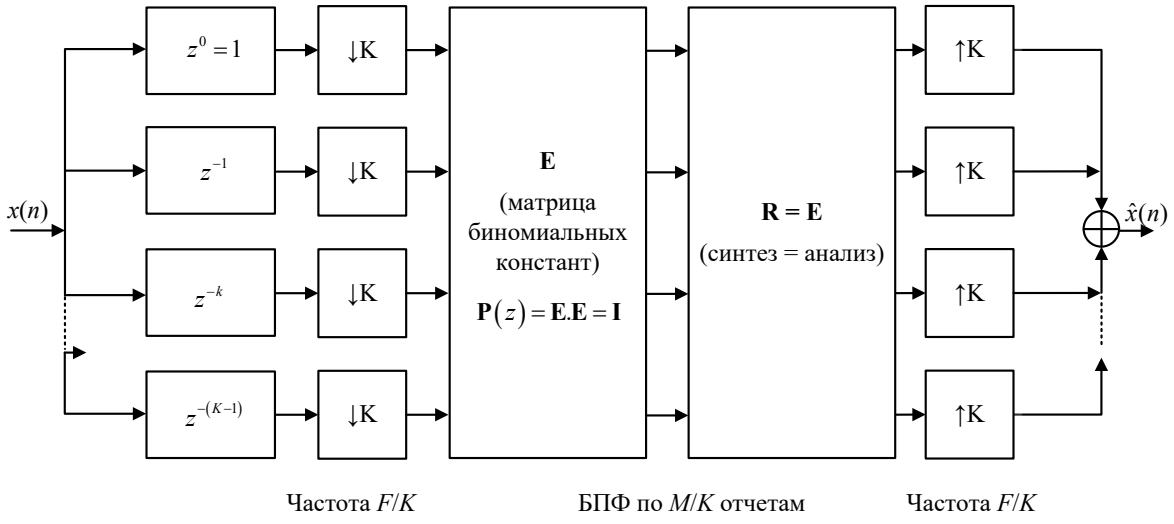


Рисунок 2.3 – Полифазная реализация банка фильтров системы на основе представления Хургина – Яковлева

2.2.3 Оценка вычислительной сложности БПФ

Рассмотрим вычислительную сложность цифровой системы передачи и обработки сигналов на основе представления Хургина – Яковлева при реализации с использованием структуры банков фильтров. Представления Хургина – Яковлева позволяет восстанавливать сигнал путем объединения информации из отсчетов прореженного сигнала и производных, а не только на основе традиционных амплитудных отсчетов. Этот подход использует гладкость сигнала во временной области, что позволяет значительно снизить частоту дискретизации. Благодаря такому снижению частоты дискретизации уменьшается объем данных, подлежащих обработке, что напрямую приводит к снижению вычислительной сложности в системах цифровой обработки. Для оценки эффективности проведем сравнение вычислительной сложности системы, основанной на представлении Хургина – Яковлева, с классической системой на основе теоремы В. А. Котельникова. Процедура оценки включает вычисление БПФ с использованием алгоритма Кули-Тьюки – стандартного метода для спектрального анализа сигналов в цифровой обработке [86]. БПФ применяется здесь как инструмент для сравнения объема вычислений, а не как часть самого представления Хургина – Яковлева.

Если используется БПФ, то количество вычислительных операций системы передачи и обработки сигналов на основе теоремы В. А. Котельникова зависит от количества отсчетов в кадре обработки M, а именно, число операций определяется по формуле:

$$C_{КОТ} = \mathrm{O}\left(M \log_2 M\right). \tag{2.9}$$

В отличие от этого, представление Хургина – Яковлева использует отсчеты производной, что позволяет значительно снизить частоту дискретизации и, соответственно, уменьшить количество отсчетов M. При использовании параллельной K-канальной системы передачи и обработки в каждом канале частота дискретизации уменьшается в K раз, по сравнению с частотой дискретизации, заданной теоремой В. А. Котельникова. В результате, количество отсчетов, необходимое для каждого канала, составляет M/K. При реализации систем обработки по предложенному алгоритму выполняется K процедур БПФ для M/K отдельных

децимированных отсчетов сигнала и его $(K-1)$ производных. Таким образом, общее число операций системы на основе представления Хургина – Яковлева в K-канальной системе обработки определяется следующим выражением:

$$C_{ХЯ_1} = \mathrm{O}\left(\underbrace{\frac{M}{K}\log_2\left(\frac{M}{K}\right)+\frac{M}{K}\log_2\left(\frac{M}{K}\right)+\ldots+\frac{M}{K}\log_2\left(\frac{M}{K}\right)}_{K\ \text{элементов}}\right) = \mathrm{O}\left(M\log_2\left(\frac{M}{K}\right)\right). \quad (2.10)$$

Оценку относительного сокращения количества вычислительных операций при использовании БПФ Кули-Тьюки в системах на основе представления Хургина – Яковлева по сравнению с системами, основанными на теореме В. А. Котельникова, можно произвести по следующей формуле:

$$\alpha_K = \frac{C_{КОТ}}{C_{ХЯ_1}} = \frac{M\log_2 M}{M\log_2\left(\frac{M}{K}\right)} = \log_{\frac{M}{K}} M. \quad (2.11)$$

На рисунке 2.4 представлены графики зависимостей относительного сокращения количества вычислительных операций α_K от количества используемых отсчетов M для различного числа каналов обработки $K = 2$ (обозначено цифрой 1), $K = 3$ (обозначено цифрой 2) и $K = 4$ (обозначено цифрой 3).

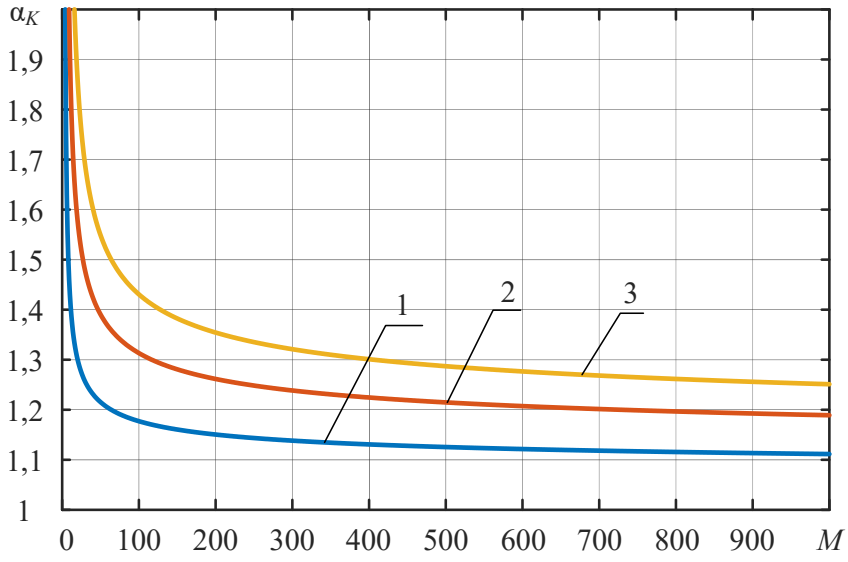


Рисунок 2.4 – Относительное сокращение количества вычислительных операций системы, при реализации БПФ на основе представления Хургина – Яковлева, для случая с количеством каналов $K = 2...4$

Анализ данных показывает, что представление Хургина – Яковлева

обеспечивает значительное сокращение количества вычислительных операции, по сравнению с системой на основе теоремы В. А. Котельникова, особенно при небольших и средних значениях числа отсчетов в блоке анализа *M*. При увеличении количества отсчетов *M* эффективность сокращения количества вычислительных операций постепенно уменьшается, и значения сокращения количества вычислительных операций α_K стремятся к 1, что указывает на снижение преимуществ алгоритма при очень больших объемах данных. Кроме того, увеличение числа параллельных каналов *K* приводит к более выраженному эффекту: наибольшее сокращение количества операций наблюдается при четырехканальной системе $K = 4$, тогда как при двухканальной системе $K = 2$ эффективность минимальна, но все же остается выше единицы. Таким образом, представление Хургина – Яковлева демонстрирует высокую эффективность в системах с ограниченным количеством отсчетов *M* и при использовании большего числа параллельных каналов обработки *K*, что делает его особенно полезным для приложений, требующих оптимизации вычислительных ресурсов.

Далее рассмотрим случай использования многоэтапной древовидной структуры, где число параллельных каналов обработки увеличивается с каждым уровнем. В случае *K*-канальной системы передачи и обработки сигналов число этапов разложения равно *P*, а общее количество используемых параллельных каналов на последнем уровне составляет K^P. В такой многоэтапной системе каждая ветвь дерева представляет собой отдельный подканал, который обрабатывает часть сигнала с уменьшенной частотой дискретизации. Это позволяет значительно снизить вычислительную сложность по сравнению с классической системой на основе теоремы В.А. Котельникова. При реализации многоэтапной структуры с K^P каналами, общая вычислительная сложность может быть оценена следующим образом: на последнем этапе *P* каждый канал обрабатывает M/K^P отсчетов, поэтому количество вычислительных операций БПФ для одного канала составляет:

$$\mathrm{O}\left(\frac{M}{K^P}\log_2\left(\frac{M}{K^P}\right)\right). \qquad (2.12)$$

Так как на последнем этапе имеется K^P каналов, общее количество вычислительных операций для всех каналов этапа P будет определяться по следующей формуле:

$$C_{ХЯ_P} = \mathrm{O}\left(\underbrace{\frac{M}{K^P}\log_2\left(\frac{M}{K^P}\right) + \frac{M}{K^P}\log_2\left(\frac{M}{K^P}\right) + \ldots + \frac{M}{K^P}\log_2\left(\frac{M}{K^P}\right)}_{K^P \text{ элементов}} \right) = \\ = \mathrm{O}\left(M\log_2\left(\frac{M}{K^P}\right)\right). \tag{2.13}$$

Относительное сокращение вычислительных затрат, при использовании БПФ Кули-Тьюки, в системах на основе представления Хургина – Яковлева, по сравнению с системами, основанными на теореме В.А. Котельникова, для системы с K каналами и P этапами может быть оценено по следующей формуле:

$$\alpha_{K^P} = \frac{C_{КОТ}}{C_{ХЯ_P}} = \frac{M\log M}{M\log\left(\frac{M}{K^P}\right)} = \log_{\frac{M}{K^P}} M. \tag{2.14}$$

На рисунках 2.5 и 2.6 представлены графики зависимостей относительного сокращения количества вычислительных операций в зависимости от числа отсчетов M для систем с различным количеством параллельных каналов K. Для обоих рисунков рассматриваются случаи $K = 2$ (обозначено цифрой 1), $K = 3$ (обозначено цифрой 2) и $K = 4$ (обозначено цифрой 3). На рисунке 2.5 представлены результаты для этапа разложения $P = 2$, а на рисунке 2.6 — для этапа разложения $P = 3$. Анализ приведенных зависимостей показывает, что представление Хургина – Яковлева обеспечивает значительное снижение вычислительной сложности, по сравнению с классической системой, основанной на теореме В.А. Котельникова, особенно при небольших значениях отсчетов M.

При увеличении числа каналов K эффективность снижения количества вычислительных операций возрастает: наибольший эффект наблюдается для $K = 4$, тогда как для $K = 2$ снижение менее выражено, но все же остается выше единицы. Кроме того, увеличение числа этапов разложения P также способствует улучшению эффективности системы на основе представления Хургина – Яковлева.

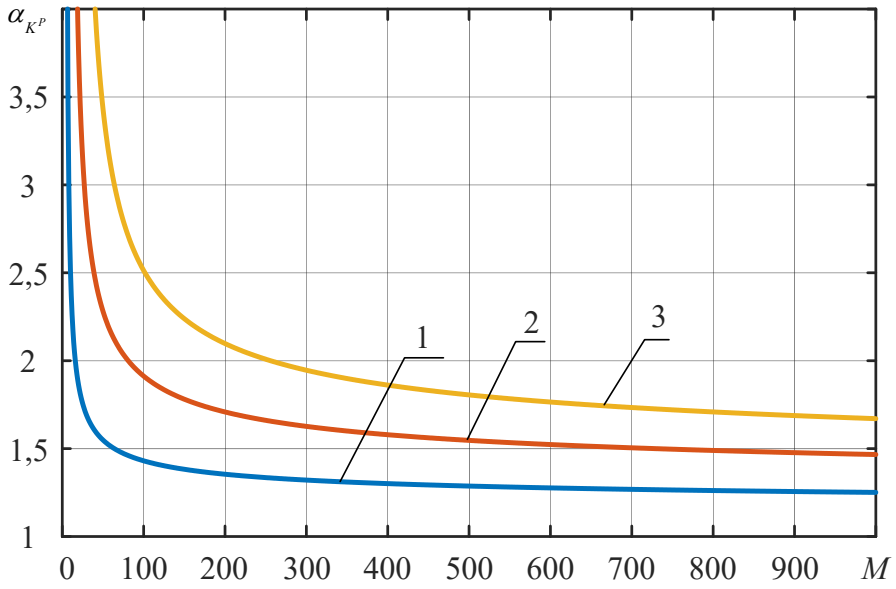


Рисунок 2.5 – Относительное сокращение количества вычислительных операций системы при реализации БПФ на основе представления Хургина – Яковлева для случая с количеством каналов $K = 2...4$ и числа этапов разложения $P = 2$

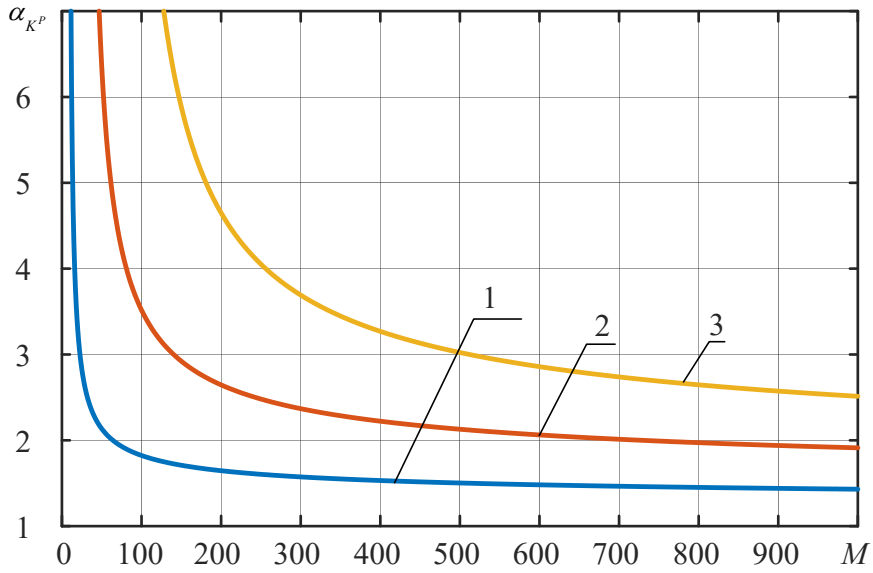


Рисунок 2.6 – Относительное сокращение количества вычислительных операций системы при реализации БПФ на основе представления Хургина – Яковлева для случая с количеством каналов $K = 2...4$ и числа этапов разложения $P = 3$

Например, при $P = 3$ (рисунок 2.6), снижение количества вычислительных операций более значительно, чем при $P = 2$ (рисунок 2.5), особенно для малых значений M. Это подтверждает, что многоэтапная структура позволяет лучше использовать преимущества уменьшения частоты дискретизации и параллельной обработки данных.

Таким образом, комбинация большого числа каналов K и этапов разложения P обеспечивает максимальную оптимизацию вычислительных ресурсов. Представление Хургина – Яковлева особенно эффективно в задачах с ограниченным количеством отсчетов M, где его преимущество перед классической системой наиболее заметно. Однако при больших значениях количество отсчетов в кадре M эффект снижения количества вычислительных операций постепенно уменьшается, хотя алгоритм все равно сохраняет свои преимущества. Эти результаты подчеркивают важность выбора параметров K и P для достижения наилучшей производительности в современных системах цифровой обработки сигналов.

Для иллюстрации выводов, представленных выше, рассмотрим применение многоканальной и многоэтапной структуры декомпозиции на основе представления Хургина – Яковлева в конкретной системе — системе передачи и обработки РС с частотой дискретизации 8 кГц. В таких системах ключевые параметры, влияющие на вычислительную эффективность, включают размер кадра анализа (определяющий количество обрабатываемых отсчетов M), количество параллельных каналов K и число этапов разложения P. Выбор временных интервалов анализа $M = 128$ и $M = 256$ отсчетов, соответствующих 16 мс и 32 мс, согласуется с длительностью квазистационарных участков речевого сигнала, что подтверждается исследованиями в области цифровой обработки сигналов.

На основе анализа приведенных зависимостей, представленных на рисунках 2.4, 2.5 и 2.6, можно оценить вычислительную сложность для систем на основе представления Хургина – Яковлева с различными значениями K и P.

В таблице 2.1 показано относительное снижение количества вычислительных операций, при использовании БПФ Кули-Тьюки в системах на основе представления Хургина – Яковлева, по сравнению с классическими системами, основанными на теореме В.А. Котельникова. Результаты демонстрируют, что в системе с двумя каналами $\left(K = 2\right)$ и числом этапов разложения $P = 1\ldots3$ относительное сокращение вычислительных затрат достигает диапазона от 1,143 до 1,75 раз для кадров анализа

размером $M = 128$ и $M = 256$. Это означает, что системы на основе представления Хургина – Яковлева позволяют обрабатывать речевые сигналы в реальном времени с существенно меньшими вычислительными затратами.

Таблица 2.1. Сокращение количества вычислительных операций системы на основе представления Хургина – Яковлева при реализации БПФ

	Общее количество каналов системы	Сокращение количества вычислительных операций				
		M=64	*M*=128	*M*=256	*M*=512	*M*=1024
K=2, *P*=1	2	1,2	1,167	1,143	1,125	1,111
K=3, *P*=1	3	1,36	1,293	1,247	1,214	1,188
K=4, *P*=1	4	1,5	1,4	1,333	1,286	1,25
K=2, *P*=2	4	1,5	1,4	1,333	1,286	1,25
K=3, *P*=2	9	2,12	1,828	1,656	1,544	1,464
K=4, *P*=2	16	3	2,333	2	1,8	1,667
K=2, *P*=3	8	2	1,75	1,6	1,5	1,429
K=3, *P*=3	27	4,82	3,118	2,465	2,12	1,907
K=4, *P*=3	64	-	7	4	3	2,5

Кроме того, использование отсчетов производных в представлении Хургина-Яковлева позволяет уменьшить влияние шумов и повысить точность восстановления сигнала, что является критически важным для обработки РС в условиях ограниченной полосы пропускания. В отличие от классических методов и алгоритмов, основанных на теореме В.А. Котельникова, где требуется высокая частота дискретизации, представление Хургина – Яковлева обеспечивает более эффективное использование вычислительных ресурсов без потери качества обработки [106, 107]. Это достигается за счет снижения частоты дискретизации в каждом канале и применения многоуровневой структуры, что позволяет значительно снизить общую вычислительную нагрузку.

Таким образом, применение многоэтапной структуры на основе представления Хургина – Яковлева в системах обработки РС не только снижает вычислительную сложность, но и повышает общую эффективность системы. Это делает его особенно подходящим для современных задач цифровой обработки речи,

где требуется оптимизация вычислительных ресурсов при сохранении высокого качества обработки.

2.3 Помехоустойчивость системы обработки речевых сигналов на основе представления Хургина – Яковлева

2.3.1 Постановка задачи и модель помех

Практическая ценность системы обработки и передачи РС на основе представления Хургина – Яковлева определяется не только точностью восстановления в идеальном канале, но и устойчивостью к реальным воздействиям, неизбежно присутствующим в каналах цифровой передачи данных. К наиболее распространённому классу таких воздействий относится аддитивный белый гауссовский шум (АБГШ), обусловленный тепловыми флуктуациями и квантовыми эффектами в приёмо-передающем оборудовании. Настоящий раздел посвящён теоретическому анализу и экспериментальным исследованиям помехоустойчивости системы обработки РС, реализованной на основе представления Хургина – Яковлева, при воздействии АБГШ в канале связи, в сравнении с системой на основе теоремы В.А. Котельникова. Рассмотрим обобщённую структурную схему системы при числе каналов K, показанную на рисунке 2.7.

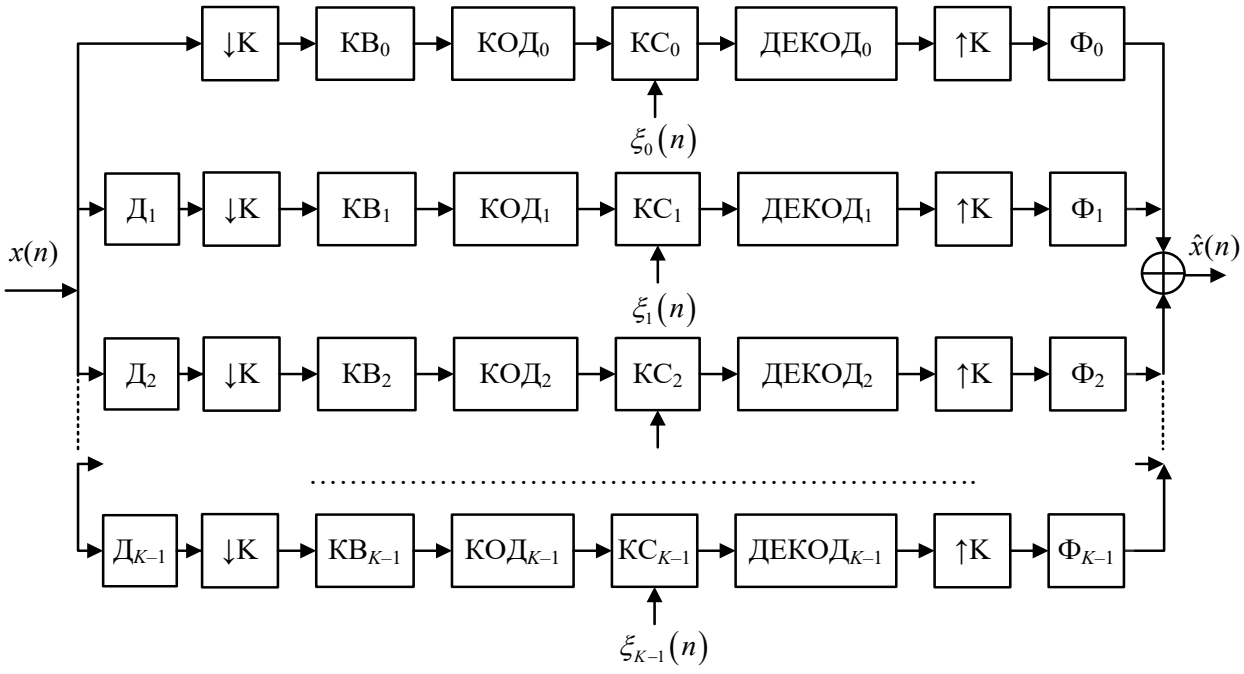


Рисунок 2.7 – Обобщённая структурная схема системы обработки РС при воздействии шума

На передающей стороне входной речевой сигнал $x(n)$ поступает на дифференциатор для получения производных порядков $1,2,\ldots,K-1$. Прореженные отсчёты сигнала и его производных передаются по K параллельным каналам, где подвергаются квантованию и кодированию. В k-м канале $(k=0,1,\ldots,K-1)$ к полезному сигналу $S_k(n)$ прибавляется аддитивный белый гауссовский шум $\xi_k(n)$ со спектральной плотностью мощности (СПМ) G_ξ, равномерной в полосе частот $[0;F]$. Предполагается, что реализации шума в различных каналах статистически независимы и одинаково распределены. На приёмной стороне декодированные отсчёты через блоки интерполяции поступают на K синтезирующих фильтров $\Phi_0,\Phi_1,\ldots,\Phi_{K-1}$, выходные сигналы которых суммируются для восстановления речевого сигнала $\hat{x}(n)$. При наличии АБГШ в каналах связи спектр восстановленного сигнала на выходе сумматора имеет вид:

$$\hat{X}(n)=\sum_{k=0}^{K-1}S_k(n)e^{j\pi k/2}A_k(n)e^{-j\pi k/2}+\sum_{k=0}^{K-1}\xi_k(n)A_k(n)\,e^{-j\pi k/2},\ 0\le n\le M-1, \qquad (2.15)$$

где $S_k(n)$ – спектр полезного сигнала k-го канала (сигнал или его производная), $\xi_k(n)$ – спектр реализации шума в k-м канале связи, $A_k(n)$ – АЧХ k-го синтезирующего фильтра; $e^{-j\pi k/2}$ – ФЧХ синтезирующего фильтра, обусловленная фазово-частотной характеристикой производной k-го порядка; M – число отсчётов в обрабатываемом кадре.

Из выражения (2.15) следует принципиальное различие в поведении полезного сигнала и шума при суммировании. Первое слагаемое представляет собой когерентную сумму: фазовая метка $e^{j\pi k/2}$, внесённая при вычислении производных на передающей стороне, точно компенсируется фазовой характеристикой $e^{-j\pi k/2}$ соответствующего синтезирующего фильтра, в результате чего сигнальные составляющие из всех K каналов складываются синфазно:

$$\sum_{k=0}^{K-1}S_k(n)e^{j\pi k/2}A_k(n)e^{-j\pi k/2}=\sum_{k=0}^{K-1}S_k(n)A_k(n)=X(n). \qquad (2.16)$$

Второе слагаемое в (2.15) описывает суммарный шум на выходе системы.

Шумовые составляющие $\xi_k(n)$ не несут фазовой метки передающей стороны, поэтому на выходе k-го синтезирующего фильтра они приобретают дополнительный фазовый сдвиг $-\pi k/2$. При числе каналов K фазовые сдвиги шумовых компонент составляют $0,-\pi/2,-\pi,\ldots,(K-1)\pi/2,$ что обусловливает их некогерентное сложение на выходе сумматора в следующем виде:

$$\Xi_{\text{вых}}(n)=\sum_{k=0}^{K-1}\xi_k(n)\,A_k(n)\,e^{-j\pi k/2}. \tag{2.17}$$

Именно это принципиальное различие между когерентным накоплением полезного сигнала и некогерентным сложением шума обусловливает выигрыш в помехоустойчивости системы на основе представления Хургина – Яковлева по отношению к классической системе на основе теоремы В.А. Котельникова, в которой передаётся единственный поток с полной дисперсией шума канала на выходе. Для количественной оценки этого выигрыша необходимо вычислить суммарную СПМ шума с учётом АЧХ и ФЧХ соотношений синтезирующих фильтров. Данный анализ проводится раздельно для случаев в разделе 2.3.2.

2.3.2 Теоретическая оценка помехоустойчивости при воздействии белого гауссовского шума

Для количественной оценки выигрыша в помехоустойчивости системы на основе представления Хургина – Яковлева необходимо вычислить суммарную мощность шума на выходе сумматора синтезирующих фильтров и сравнить её с мощностью шума в эталонной системе на основе теоремы В.А. Котельникова при одинаковых условиях передачи. В системе на основе теоремы В.А. Котельникова передаётся единственный поток отсчётов сигнала. Шум канала $\xi_k(n)$ с равномерной СПМ G_ξ проходит через синтезирующий фильтр с единичной АЧХ, поэтому средняя мощность шума на выходе имеет вид:

$$P_{\xi_{КОТ}}=\int_0^F G_\xi df = G_\xi F. \tag{2.18}$$

В системе на основе представления Хургина – Яковлева при числе каналов K шум воздействует на K параллельных каналов. Для обеспечения корректности

сравнения предполагается, что суммарная мощность передаваемых данных в обеих системах одинакова. Это достигается за счёт нормировки уровней сигналов в каналах. Условие нормировки: средняя мощность отсчётов сигнала и производных в каждом канале совпадает с мощностью отсчётов исходного сигнала в базовой системе. При этом шум $\xi_k(n)$ в каждом из K каналов статистически независим и имеет ту же СПМ G_ξ.

Для справедливого сравнения с системой, основанной на представлении Хургина – Яковлева при K каналах рассмотрим эквивалентную систему В.А. Котельникова, передающую те же K потоков, — её суммарная мощность шума составляет:

$$P_{\xi_{KOT}}^{(K)} = KG_\xi F. \tag{2.19}$$

Из выражения (2.17) следует, что при статистической независимости шумов в каналах суммарная мощность шума на выходе определяется модулем результирующего фазора. Введём энергетический коэффициент k-го синтезирующего фильтра как интеграл от его АЧХ по рабочей полосе:

$$K_{\varphi_k} = \int_0^F A_k(f)\,df. \tag{2.20}$$

Тогда мощность суммарного выходного шума:

$$P_{\xi_{XЯ}}^{(K)} = G_\xi \left| \sum_{k=0}^{K-1} K_{\varphi_k} e^{-j\pi k/2} \right|. \tag{2.21}$$

Физический смысл формулы (2.21): K_{φ_k} задаёт амплитуду шумового фазора k-го канала, множитель $e^{-j\pi k/2}$ – его ориентацию в комплексной плоскости. Чем сильнее взаимная некогерентность фазоров (чем ближе их ориентации к взаимно противоположным), тем меньше модуль результирующего вектора и, следовательно, тем меньше мощность шума на выходе системы передачи.

Выигрыш в ОСШ:

$$Q_K = \frac{P_{\xi_{KOT}}^{(K)}}{P_{\xi_{XЯ}}^{(K)}} = \frac{KF}{\left| \sum_{k=0}^{K-1} K_{\varphi_k} e^{-j\pi k/2} \right|}. \tag{2.22}$$

Формула (2.22) является ключевым результатом теоретического анализа. Она показывает, что выигрыш определяется двумя факторами: числом каналов K и степенью некогерентности шумовых фазоров. С ростом K числитель линейно возрастает, тогда как знаменатель возрастает значительно медленнее ввиду взаимной компенсации фазоров — именно это и обеспечивает монотонный рост Q_K с увеличением числа каналов.

Перейдём к вычислению конкретных значений выигрыша для $K = 2,3,4$.

Рассмотрим двухканальную систему $(K = 2)$ на основе представления Хургина – Яковлева. АЧХ синтезирующих фильтров Φ_{20} и Φ_{21} определены выражениями (1.8), (1.9). Вычислим энергетические коэффициенты по формуле (2.20):

$$K_{\varphi_0} = \int_0^F 2\left(1 - \frac{f}{F}\right) df = F, \;\; K_{\varphi_1} = \int_0^F \frac{2}{\pi} df = \frac{2}{\pi} F. \tag{2.23}$$

При $K = 2$ фазовые сдвиги шумовых фазоров составляют 0 и $-\pi/2$ − компоненты ортогональны в комплексной плоскости, как показано на рисунке 2.8а. Разложим результирующий фазор на вещественную и мнимую части:

$$\sum_{k=0}^{1} K_{\varphi_k} e^{-j\pi k/2} = K_{\varphi_0} - jK_{\varphi_1} = F - j\frac{2}{\pi} F. \tag{2.24}$$

Мощность выходного шума:

$$P_{\xi_{ХЯ}}^{(2)} = G_\xi F \sqrt{1^2 + \left(\frac{2}{\pi}\right)^2} = G_\xi F \sqrt{1 + \frac{4}{\pi^2}}. \tag{2.25}$$

Выигрыш в ОСШ:

$$Q_2 = \frac{P_{\xi_{КОТ}}^{(2)}}{P_{\xi_{ХЯ}}^{(2)}} = \frac{2F}{F\sqrt{1 + \frac{4}{\pi^2}}} \approx 1,687, \;\; Q_2^{\text{дБ}} = 10\lg(1,687) \approx 2,27 \text{дБ}. \tag{2.26}$$

Рассмотрим трёхканальную систему $(K = 3)$ на основе представления Хургина – Яковлева. Энергетические коэффициенты фильтров $\Phi_{30}, \Phi_{31}, \Phi_{32},$, вычисленные по формуле (2.20) из АЧХ (1.23)…(1.25):

$$K_{\varphi_0}^{(3)} = 1,0002F, \; K_{\varphi_1}^{(3)} = 0,7163F, \; K_{\varphi_2}^{(3)} = 0,9348F. \tag{2.27}$$

Раскроем сумму (2.21) для $K=3$, подставляя фазовые сдвиги $0,-\pi/2,\ -\pi$:

$$\sum_{k=0}^{2} K_{\varphi_k}^{(3)} e^{-j\pi k/2} = K_{\varphi_0}^{(3)} - jK_{\varphi_1}^{(3)} - K_{\varphi_2}^{(3)} = \left(K_{\varphi_0}^{(3)} - K_{\varphi_2}^{(3)}\right) - jK_{\varphi_1}^{(3)}. \qquad (2.28)$$

Подставляя численные значения из (2.27):

$$\sum_{k=0}^{2} K_{\varphi_k}^{(3)} e^{-j\pi k/2} = \left(1{,}0002 - 0{,}9348\right)F - j.0{,}7163F = 0{,}0654F - j.0{,}7163F. \qquad (2.29)$$

Мощность шума на выходе системы передачи:

$$P_{\xi_{XЯ}}^{(3)} = G_\xi F\sqrt{0{,}0654^2 + 0{,}7163^2} \approx 0{,}7193 G_\xi F. \qquad (2.30)$$

Выигрыш в ОСШ для трехканальной системы на основе представления Хургина – Яковлева:

$$Q_3 = \frac{P_{\xi_{KOT}}^{(3)}}{P_{\xi_{XЯ}}^{(3)}} = \frac{3F}{0{,}7193F} \approx 4{,}1708,\ Q_3^{\text{дБ}} = 10\lg\left(4{,}1708\right) \approx 6{,}202\text{дБ}. \qquad (2.31)$$

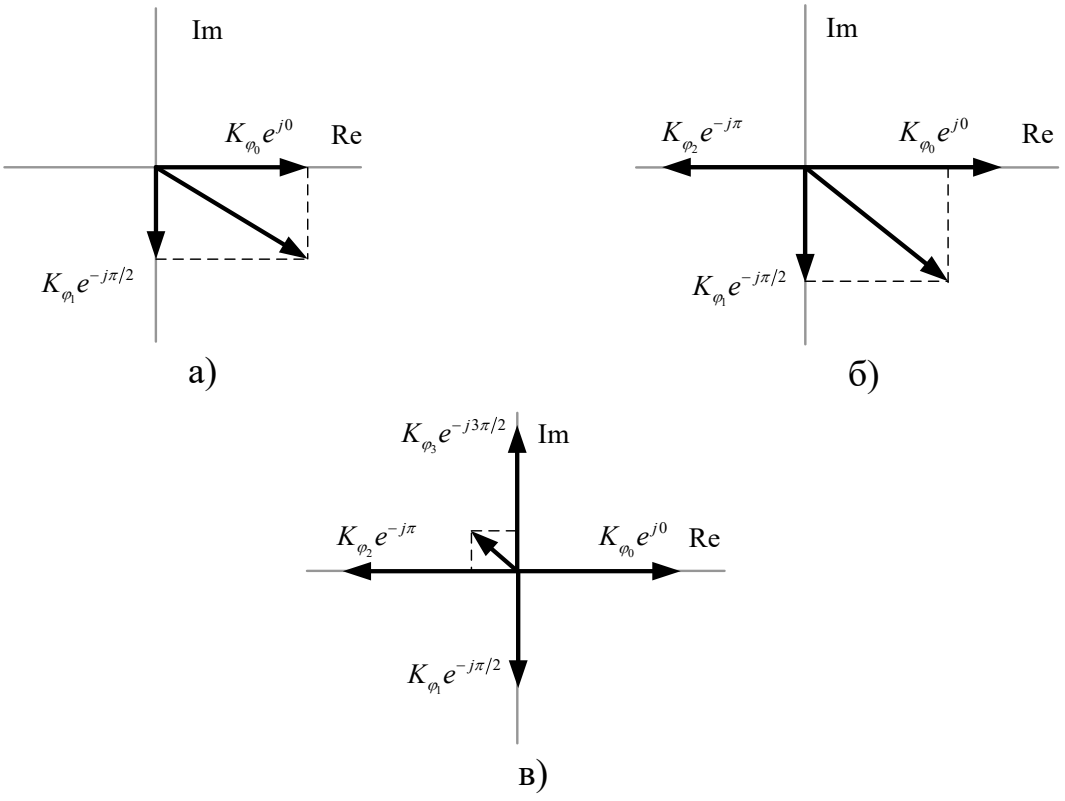


Рисунок 2.8 – Фазорные диаграммы шумовых составляющих на выходе сумматора при: а) $K=2$, б) $K=3$, в) $K=4$

Рассмотрим четырёхканальную систему $\left(K=4\right)$ на основе представления Хургина – Яковлева. Энергетические коэффициенты фильтров $\Phi_{40},\Phi_{41},\Phi_{42},\Phi_{43}$, вычисленные по формуле (2.20) из АЧХ (1.31)…(1.34):

$$K_{\varphi_0}^{(4)} = 1{,}001F,\ K_{\varphi_1}^{(4)} = 1{,}641F,\ K_{\varphi_2}^{(4)} = 1{,}345F,\ K_{\varphi_3}^{(4)} = 2{,}060F. \tag{2.32}$$

Раскроем сумму (2.21) для $K = 4$, подставляя фазовые сдвиги $0, -\pi/2$, $-\pi$, $-3\pi/2$:

$$\sum_{k=0}^{3} K_{\varphi_k}^{(4)} e^{-j\pi k/2} = K_{\varphi_0}^{(4)} - jK_{\varphi_1}^{(4)} - K_{\varphi_2}^{(4)} + jK_{\varphi_3}^{(4)} = \left(K_{\varphi_0}^{(4)} - K_{\varphi_2}^{(4)}\right) + j\left(K_{\varphi_3}^{(4)} - K_{\varphi_1}^{(4)}\right). \tag{2.33}$$

Подставляя численные значения из формулы (2.32):

$$\begin{aligned} &\sum_{k=0}^{3} K_{\varphi_k}^{(4)} e^{-j\pi k/2} = \left(1{,}001 - 1{,}345\right)F + j\left(2{,}060 - 1{,}641F\right) \\ &= -0{,}344F + j.0{,}419F. \end{aligned} \tag{2.34}$$

Мощность выходного шума:

$$P_{\xi_{ХЯ}}^{(4)} = G_\xi F\sqrt{\left(-0{,}344\right)^2 + 0{,}419^2} \approx 0{,}542 G_\xi F. \tag{2.35}$$

Выигрыш в ОСШ:

$$Q_4 = \frac{P_{\xi_{КОТ}}^{(4)}}{P_{\xi_{ХЯ}}^{(4)}} = \frac{4F}{0{,}542F} \approx 7{,}376,\ Q_4^{\text{дБ}} = 10\lg\left(9{,}389\right) \approx 8{,}68\text{дБ}. \tag{2.36}$$

Результаты расчётов для всех трёх конфигураций систематизированы в таблице 2.2. Из таблицы 2.2 видно, что выигрыш монотонно возрастает с ростом *K*, однако темп роста неравномерен: переход от $K = 2$ к $K = 3$ даёт прирост 3,93 дБ, тогда как переход от $K = 3$ к $K = 4$ – значительно больший прирост 2,48 дБ. Рассмотрим многоэтапное разложение на основе представления Хургина – Яковлева. В разделе 2.2 было показано, что данное представление допускает *P*-этапное иерархическое разложение, при котором каждый последующий уровень обрабатывает низкочастотную составляющую выхода предыдущего.

Таблица 2.2. Теоретический выигрыш в ОСШ при одноэтапном разложении

K	$P_{\xi_{КОТ}}^{(K)}$	$P_{\xi_{ХЯ}}^{(K)}$	Q_K, раз	Q_K, дБ
2	$2G_\xi F$	$1{,}185G_\xi F$	1,687	2,27
3	$3G_\xi F$	$0{,}719G_\xi F$	4,171	6,20
4	$4G_\xi F$	$0{,}542G_\xi F$	7,376	8,68

С точки зрения помехоустойчивости это означает, что шум, действующий на

p-м уровне разложения, проходит через ту же *K*-канальную синтезирующую структуру, что и на первом уровне. Поскольку реализации шума на разных уровнях статистически независимы, каждый уровень вносит самостоятельный выигрыш Q_K. Суммарный выигрыш помехоустойчивости при *P* этапах:

$$Q_K^{(P)} = \left(Q_K\right)^P,\ Q_K^{(P)}\Big|_{\text{дБ}} = P.Q_K^{\text{дБ}}. \tag{2.37}$$

Линейный рост выигрыша в децибелах с числом этапов *P* при фиксированном *K* является важным практическим преимуществом многоэтапного разложения: каждый дополнительный уровень гарантированно добавляет фиксированный прирост помехоустойчивости $Q_K^{\text{дБ}}$ к суммарному ОСШ. Численные значения выигрыша в ОСШ приведены в таблице 2.3.

Таблица 2.3. Теоретический выигрыш в ОСШ при многоэтапном разложении, дБ

число этапов / число каналов	$P = 1$	$P = 2$	$P = 3$
$K = 2$	2,27	4,54	6,81
$K = 3$	6,20	12,40	18,61
$K = 4$	8,68	17,36	26,04

2.3.3 Экспериментальное исследование помехоустойчивости системы обработки речевых сигналов на основе представления Хургина – Яковлева

Теоретические значения выигрыша Q_K, полученные в разделе 2.3.2, характеризуют идеальный механизм подавления шума при статистически независимых шумовых составляющих в каналах. В реальной реализованной системе на выходе сумматора одновременно действуют два разнородных источника погрешности: внешний шум канала связи и собственная погрешность алгоритма $\varepsilon_\Sigma = \varepsilon_\phi + \varepsilon_{der} + \varepsilon_a$, исследованная в разделе 1.6. Поведение системы в условиях шумового канала не может быть описано ни теоретической моделью раздела 2.3.2 в отдельности, ни данными таблицы 1.1 в отдельности. Задача настоящего раздела — построить аналитическую модель, объединяющую оба источника погрешности, и верифицировать её экспериментально.

Исследования проводились для набора из 100 записей РС от 10 дикторов различного пола в возрасте от 20 до 60 лет согласно требованиям ГОСТ Р 50840-95 [73]. Параметры оцифровки: частота дискретизации $F_{Д} = 8$ кГц, разрядность 16 бит, обработка по непересекающимся кадрам длиной $M = 256$ отсчётов (32 мс). К каждому фрагменту добавлялась реализация АБГШ с заданным значением ОСШ $SNR_{вх}$ из ряда 0, 5, 10, 15, 20, 25, 30 дБ. Для каждой конфигурации (K, P) выполнялось 500 независимых испытаний; доверительный интервал при уровне значимости $\alpha = 0{,}01$ не превышал 0,3% в соответствии с нормальным законом распределения (критерий Колмогорова – Смирнова). Обработка осуществлялась на платформе MATLAB R2023b. Оценка качества производилась по двум метрикам: инструментальной ОСШ — $SNR_{вых}$ по формуле (1.1), и перцептивной — ViSQOL по шкале MOS 1…5.

На выходе сумматора системы на основе представления Хургина – Яковлева формируются два аддитивных компонента погрешности.

Первый — канальный шум, мощность которого на выходе уменьшена в Q_K раз по сравнению с входом.

Второй — собственная погрешность алгоритма ε_{Σ}, мощность которой при фиксированном (K, P) определяется исключительно точностью реализации синтезирующих фильтров и не зависит от шума канала: ε_a обусловлена полиномиальной аппроксимацией АЧХ, ε_{der} — точностью вычисления производных. Различная физическая природа этих составляющих — внешний АБГШ и детерминированные ошибки фильтрации — обосновывает их статистическую некоррелированность. При данном допущении мощности погрешностей складываются:

$$\sigma^2_{вых} = \frac{\sigma^2_{кан}}{Q_K} + \sigma^2_{рек}, \tag{2.38}$$

где $\sigma^2_{кан}$ — мощность канального шума, $\sigma^2_{рек}$ — мощность собственной погрешности алгоритма. Разделив обе части на мощность сигнала σ^2_s и обозначив

$SNR_{\text{рек}} = \sigma_s^2 / \sigma_{\text{рек}}^2$ — предельное ОСШ алгоритма при идеальном канале, получаем:

$$\frac{1}{SNR_{\text{вых}}} = \frac{1}{SNR_{\text{рек}}} + \frac{1}{Q_K . SNR_{\text{вх}}}. \tag{2.39}$$

Формула (2.39) не содержит свободных параметров: Q_K определены в разделе 2.3.2, $SNR_{\text{рек}}$ взяты непосредственно из таблицы 1.1. Она воспроизводит оба предельных режима корректно: при $SNR_{\text{вх}} \to 0$ доминирует канальный шум и $SNR_{\text{вых}} \approx Q_K$. $SNR_{\text{вх}}$ — теоретический выигрыш реализуется в полной мере; при $SNR_{\text{вх}} \to \infty$ система выходит на насыщение $SNR_{\text{вых}} \to SNR_{\text{рек}}$.

Из (2.39) следует существование порогового значения $SNR_{\text{вх}}^*$, при превышении которого $SNR_{\text{вых}} \approx Q_K$ для системы на основе представления Хургина – Яковлева оказывается ниже, чем для системы на основе теоремы В.А. Котельникова $(Q_K = 1)$. Приравнивая $SNR_{\text{вых}}^{\text{ХЯ}} = SNR_{\text{вых}}^{\text{КОТ}} = SNR_{\text{вх}}$, получаем:

$$SNR_{\text{вх}}^* = SNR_{\text{рек}} . \frac{Q_K - 1}{Q_K}. \tag{2.40}$$

Подставляя значения из таблицы 1.1 и раздела 2.3.2, находим пороговые значения, сведённые в таблицу 2.4.

Таблица 2.4. Пороговые значения $SNR_{\text{вх}}^*$

Числа каналов	Числа этапов	$SNR_{\text{рек}}$, дБ	Q_K, дБ	$SNR_{\text{вх}}^*$, дБ
$K = 2$	$P = 1$	85,38	2,27	81,48
	$P = 2$	55,93	4,54	54,05
	$P = 3$	48,88	6,81	47,86
$K = 3$	$P = 1$	26,02	6,20	24,83
	$P = 2$	24,47	12,40	24,21
	$P = 3$	24,04	18,60	23,98
$K = 4$	$P = 1$	17,92	8,68	17,29
	$P = 2$	15,66	17,36	15,58
	$P = 3$	14,03	26,04	14,02

На рисунке 2.9 представлены теоретические кривые по модели (2.39) и экспериментальные точки для $P = 1$. Анализ рисунка 2.9 подтверждает нелинейный

характер зависимостей: при малых $SNR_{вх}$ все кривые $K = 2,3,4$ параллельны линии В.А. Котельникова со смещением Q_K, то есть теоретический выигрыш раздела 2.3.2 реализуется; с ростом $SNR_{вх}$ кривые отклоняются от линейного хода и выходят на горизонтальные асимптоты, определяемые $SNR_{рек}$ из таблицы 1.1. Экспериментальные точки отклоняются от расчётных кривых не более чем на 0,3 дБ — в пределах доверительного интервала, что подтверждает адекватность допущения о некоррелированности двух источников погрешности, принятого при выводе (2.39).

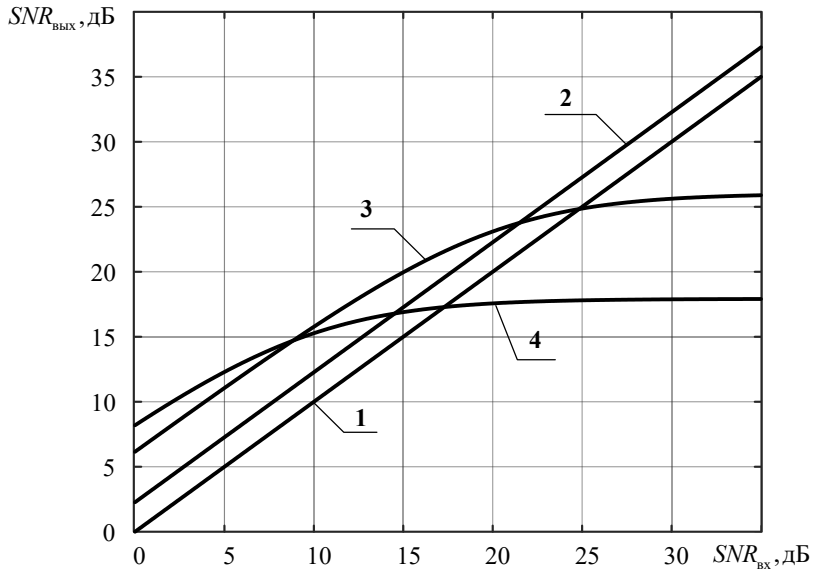


Рисунок 2.9 – Зависимости ОСШ на выходе от ОСШ на входе при действии БГШ

На рисунке 2.10 показаны зависимости перцептивной оценки качества восстановленной речи ViSQOL от уровня входной помехи при $P = 1$.

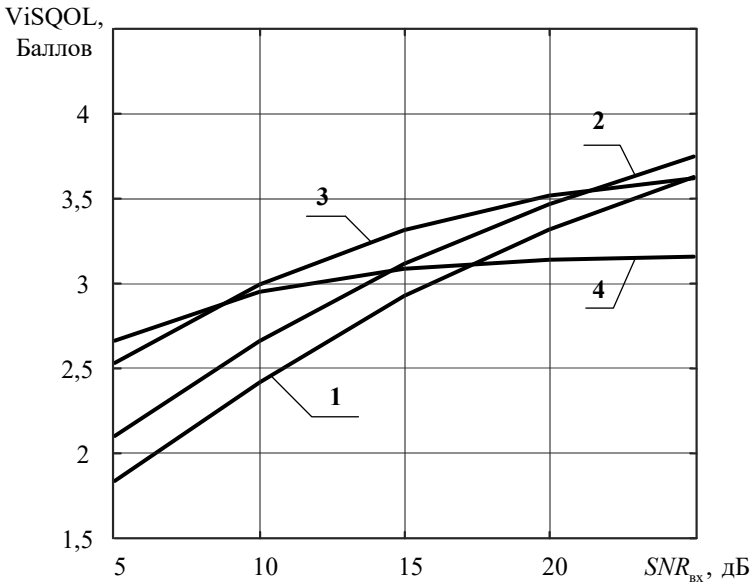


Рисунок 2.10 – Зависимости перцептивной оценки ViSQOL от ОСШ на входе

В зоне интенсивных помех ($SNR_{вх} \le 10$ дБ) все конфигурации системы, основанные на представлении Хургина – Яковлева, обеспечивают положительный выигрыш качества по критерию ViSQOL — максимум 0,81 балла при K = 4, $SNR_{вх} = 5$ дБ. При $SNR_{вх} > 17$ дБ кривая K = 4 опускается ниже уровня качества системы на основе теоремы В.А. Котельникова, тогда как K = 3 сохраняет положительный уровень вплоть до $SNR_{вх} \approx 25$ дБ.

2.4 Экспериментальное исследование эффективности алгоритмов первичного кодирования речевых сигналов

2.4.1 Методика и условия экспериментального исследования

Целью экспериментальных исследований является проверка гипотезы об эффективности предложенного алгоритма дискретизации и первичного кодирования РС на основе представления Хургина – Яковлева с многоэтапным разложением по сравнению с традиционной системой, основанной на теореме В.А. Котельникова, в условиях различных режимов кодирования и наличия ошибок в канале связи. Тестовый речевой материал формировался в соответствии с требованиями ГОСТ Р 50840–95 и рекомендаций ITU-T P.800 [72, 73]. В качестве тестовой базы использовался набор из 100 коротких речевых фраз длительностью 5…10 секунд записанных дикторами различного пола и возрастных групп от 20 до 60 лет. Параметры цифрового речевого сигнала: частота дискретизации 8 кГц, разрядность 16 бит, что соответствует стандарту телефонной полосы частот 300…3400 Гц. Все записи прошли предварительную нормализацию по уровню громкости с целью исключения влияния амплитудных различий на результаты оценки.

Для моделирования алгоритмов первичного кодирования и декодирования применялась программная среда MATLAB R2024a в сочетании с библиотекой речевых кодеков «Spirit DSP». В рамках исследования рассматривались следующие конфигурации системы представления РС: число каналов K = 2, 3, 4; число этапов разложения P = 1, 2, 3. Итого 9 конфигураций ($K \times P$) охватывали весь значимый диапазон параметров. В качестве алгоритмов первичного кодирования, совместно с которыми исследовалась предложенная система, были выбраны три кодека,

характерных для различных классов скоростей передачи:

– MMBE (Mixed-Excitation Multi-Band Excitation) со скоростью передачи 1,2 кбит/с – низкоскоростное кодирование, предназначенное для систем с жёсткими ограничениями пропускной способности, например, коротковолновых и спутниковых каналов;

– G.723.1 со скоростью передачи 6,3 кбит/с – среднескоростное кодирование, широко применяемое в системах передачи голоса по IP-сетям (VoIP) и в сотовых сетях второго поколения;

– G.726 со скоростью передачи 24 кбит/с – высокоскоростное кодирование на основе адаптивной дифференциальной импульсно-кодовой модуляции (АДИКМ), обеспечивающее высокое качество при умеренной скорости передачи.

Выбор указанных трёх кодеков обусловлен необходимостью охватить диапазон скоростей передачи, типичных для систем радиосвязи: от низкоскоростных тактических каналов до широкополосных инфраструктурных сетей, что позволяет оценить универсальность предложенного метода.

Для объективной оценки качества восстановленной речи в приёмнике применялся комплекс метрик, сформированный в разделе 1.2.2:

– SNR (Signal-to-Noise Ratio) – отношение сигнал/шум в децибелах, характеризующее интегральный уровень погрешности восстановления;

– MSE (Mean Squared Error) – среднеквадратическая ошибка, отражающая абсолютную энергию погрешности восстановления;

– ViSQOL (Virtual Speech Quality Objective Listener) – перцептуальная метрика оценки качества речи по шкале MOS (Mean Opinion Score) в диапазоне 1,0…5,0 баллов, разработанная в «Google Research», обеспечивающая коэффициент корреляции с субъективными оценками слушателей не менее 0,9. Метрика учитывает психоакустические особенности восприятия речи: маскировку, характеристики критических полос слуха и нейросетевую модель внутреннего представления звука.

Пороговые значения ViSQOL: ≥ 4,0 балла – высокое качество по категории «хорошо» согласно ITU-T P.800; ≥ 3,5 балла – приемлемое качество для систем

телефонной связи; < 3,0 балла – неприемлемое качество. Поскольку традиционные энергетические метрики SNR и MSE не всегда адекватно отражают субъективное восприятие, метрика ViSQOL принята в качестве основного критерия при сравнительном анализе эффективности анализируемых систем.

Экспериментальное исследование разделено на два последовательных этапа, каждый из которых решает самостоятельную задачу и опирается на результаты предыдущего: на первом этапе проводится оценка выигрыша качества восстановленной речи при первичном кодировании в условиях отсутствия ошибок канала связи, что позволяет определить базовую эффективность алгоритма сжатия речевого сигнала. На втором этапе проводится оценка помехоустойчивости системы в условиях моделирования канала связи с битовыми ошибками, что подтверждает работоспособность и преимущества разработанного метода при эксплуатации в реальных условиях.

2.4.2 Оценка выигрыша качества восстановленной речи в условиях первичного кодирования

На первом этапе исследовалась эффективность предложенного метода в составе предложенной системы первичного кодирования при передаче по идеальному каналу (без ошибок). Данный этап непосредственно связывает алгоритмическую точность, установленную в предыдущем разделе, с практической ценностью предложенного метода при совместном использовании с реальными кодеками. Ключевым показателем является выигрыш в качестве восстановленной речи:

$$\Delta Q = Q_{ХЯ} - Q_{КОТ}, \tag{2.41}$$

где $Q_{ХЯ}$ – оценка ViSQOL системы, реализованной на основе представления Хургина – Яковлева, $Q_{КОТ}$ – оценка ViSQOL эталонной системы, построенной на основе теоремы В.А. Котельникова при той же скорости передачи и том же алгоритме первичного кодирования.

Перед вычислением выигрыша ΔQ необходимо установить абсолютные значения качества восстановленной речи для каждой из двух сравниваемых систем.

В таблице 2.5 приведены значения метрики ViSQOL (баллов по шкале MOS) для системы на основе представления Хургина – Яковлева при всех 9 конфигурациях (K = 2, 3, 4; P = 1, 2, 3) и для эталонной системы, основанной на теореме В.А. Котельникова, при трёх кодеках первичного кодирования. Эталонная система соответствует одноканальной обработке (K = 1) при той же скорости передачи и идентичном алгоритме кодирования.

Таблица 2.5. Значения ViSQOL (баллов по шкале MOS) для системы на основе представления Хургина – Яковлева и эталонной системы на основе теоремы В.А. Котельникова при различных конфигурациях разложения и алгоритмах первичного кодирования

Числа каналов	Числа этапов	Алгоритм первичного кодирования		
		MMBE (1,2 кбит/с)	G.723.1 (6,3 кбит/с)	G.726 (24 кбит/с)
K = 1		3,72	3,92	4,21
K = 2	P = 1	4,01	4,11	4,57
	P = 2	4,78	4,59	4,71
	P = 3	4,55	4,65	4,72
K = 3	P = 1	4,28	4,53	4,40
	P = 2	4,58	4,67	4,70
	P = 3	4,62	4,68	4,70
K = 4	P = 1	4,44	4,61	4,67
	P = 2	4,57	4,67	4,71
	P = 3	4,64	4,69	4,71

На основании абсолютных значений качества вычисляется выигрыш ΔQ. На рисунке 2.11 представлены графики зависимостей выигрыша в качестве восстановленной речи ΔQ от количества этапов P, полученные при применении следующих кодеков: MMBE со скоростью передачи 1,2 кбит/с, G.723.1 со скоростью передачи 6,3 кбит/с и G.726 со скоростью передачи 24 кбит/с в системах с различными этапами разложения (P = 1, 2, 3). Под цифрой 1 указаны двухканальные системы (K = 2), под цифрой 2 — трехканальные системы (K = 3), а под цифрой 3 — четырехканальные системы (K = 4).

Как видно из приведенных зависимостей, выигрыш в качестве восстановления речи для систем на основе представления Хургина – Яковлева ΔQ остается положительными во всех рассмотренных случаях. Положительное значение ΔQ означает, что система на основе представления Хургина – Яковлева обеспечивает более высокое воспринимаемое качество восстановленной речи по сравнению с классической системой при эквивалентных условиях передачи.

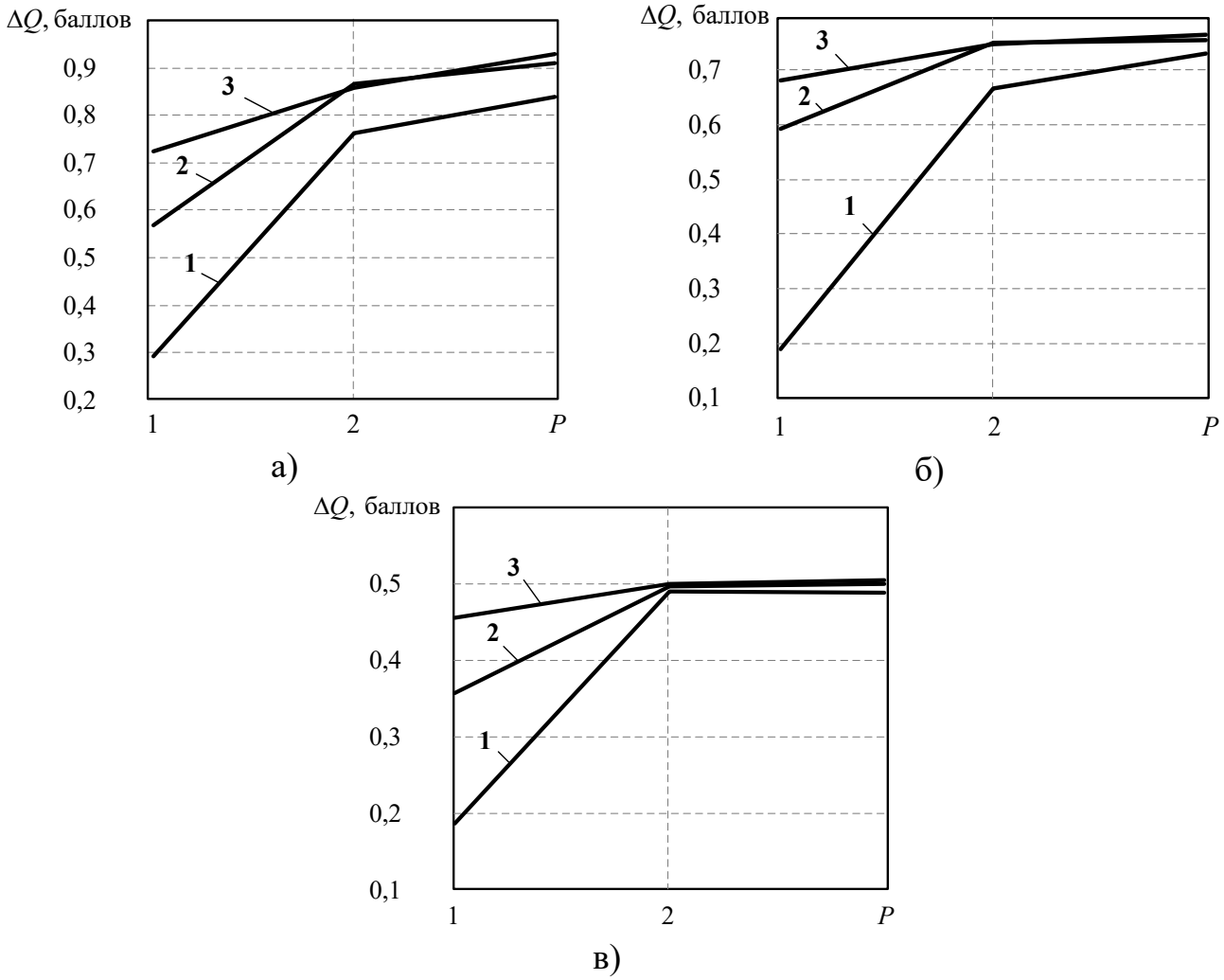


Рисунок 2.11 – Выигрыш в качестве восстановления РС при различных K и P:

а) MMBE (1,2 кбит/с); б) G.723.1 (6,3 кбит/с); в) G.726 (24 кбит/с)]

Для низкоскоростного кодека MMBE (1,2 кбит/с) система на основе представления Хургина – Яковлева демонстрирует наибольший выигрыш в диапазоне 0,29…0,92 балла по шкале MOS. Значительность выигрыша объясняется тем, что при столь низкой скорости передачи алгоритм MMBE вынужден выполнять агрессивное параметрическое сжатие, которое особенно чувствительно к шумам и ошибкам, возникающим при первичном кодировании речи.

Для кодека G.723.1 (6,3 кбит/с) выигрыш ΔQ наблюдается во всём диапазоне конфигураций, что подтверждает устойчивость предложенного метода при среднескоростном кодировании. При переходе от одноэтапного к трёхэтапному разложению выигрыш возрастает, что указывает на синергетическое взаимодействие глубины разложения с алгоритмом кодирования.

Для высокоскоростного кодека G.726 (24 кбит/с) выигрыш ΔQ хотя и меньше, чем для низкоскоростных кодеков, остаётся положительным для всех рассмотренных конфигураций. Это указывает на то, что даже при высоких скоростях передачи, когда алгоритм первичного кодирования вносит меньшие искажения, предложенный метод сохраняет преимущество. Данный результат важен для широкого спектра практических применений, поскольку подтверждает применимость метода без ограничений по классу используемого кодека.

Установлено, что с ростом числа этапов P и числа каналов K выигрыш ΔQ закономерно возрастает для всех трёх кодеков. Это согласуется с теоретическим анализом раздела 2.3, согласно которому увеличение числа каналов K обеспечивает большую глубину разложения и точнее воспроизводит форму сигнала на приёмной стороне, а увеличение числа этапов P дополнительно усиливает этот эффект за счёт многократного применения разложения.

2.4.3 Сравнительный анализ помехоустойчивости предложенного метода и системы на основе теоремы В.А. Котельникова

Далее проведено исследование помехоустойчивость системы в условиях наличия ошибок в канале связи. Ошибки в канале являются ключевым фактором деградации качества речи в реальных системах радиосвязи, особенно в условиях замираний и помех. Для моделирования ошибок в канале применялось программное обеспечение «Vocoder», обеспечивающее формирование случайных ошибок с вероятностью $P_{ош}$ в диапазоне 0,1…5 %. Данный диапазон охватывает типичные условия работы радиоканалов: от нормальных $\left(P_{ош} = 0\%\right)$ до сложной помеховой обстановке $\left(P_{ош} = 5\%\right)$.

Физический механизм выигрыша в помехоустойчивости системы на основе

представления Хургина – Яковлева обусловлен следующим. Синтезирующие фильтры системы вносят дополнительный фазовый сдвиг от 0 до $3\pi/2$ с шагом $\pi/2$ для отсчётов сигнала и его производных различных порядков. При суммировании восстановленных частичных сигналов в суммирующем устройстве приёмника составляющие ошибок, действующих в различных каналах обработки, складываются геометрически с различными фазовыми сдвигами, что обеспечивает частичную взаимную компенсацию шумов. Степень компенсации нарастает с увеличением числа каналов K вследствие снижения корреляции ошибок между каналами.

Сравнительный анализ помехоустойчивости начинается с представления абсолютных значений качества восстановленной речи в зависимости от вероятности ошибок в канале $P_{ош}$ для обеих сравниваемых систем. На рисунке 2.12 представлены зависимости качества восстановленной речи $Q_{КОТ}$ (ViSQOL, баллов по шкале MOS) от вероятности ошибок в канале $P_{ош}$ для эталонной системы на основе теоремы В.А. Котельникова при трёх исследованных алгоритмах первичного кодирования.

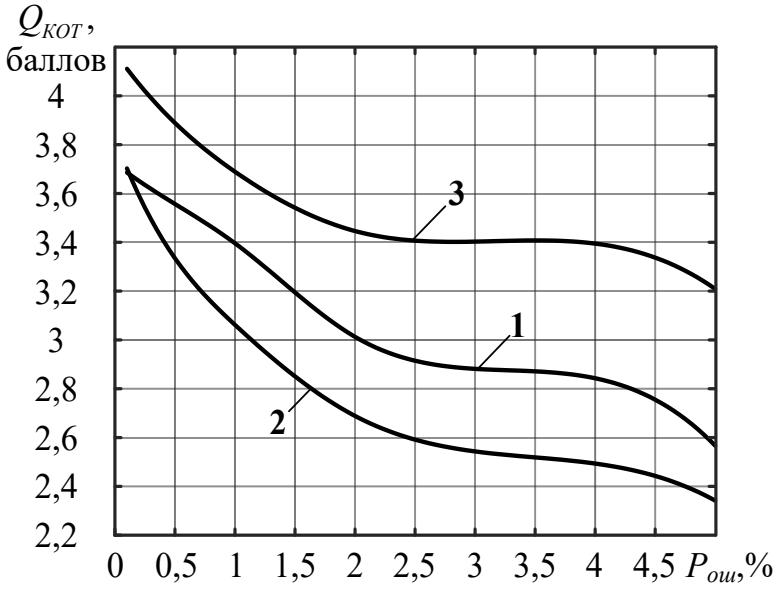


Рисунок 2.12 – Зависимость качества восстановленной речи $Q_{КОТ}$ от вероятности ошибок в канале $P_{ош}$ для системы на основе теоремы В.А. Котельникова:

1 — MMBE (1,2 кбит/с); 2 — G.723.1 (6,3 кбит/с); 3 — G.726 (24 кбит/с)]

На рисунке 2.13 представлены аналогичные зависимости $Q_{ХЯ}$ от $P_{ош}$ для системы на основе представления Хургина – Яковлева при всех исследованных конфигурациях ($K = 2, 3, 4$; $P = 1, 2, 3$) и тех же алгоритмах первичного кодирования.

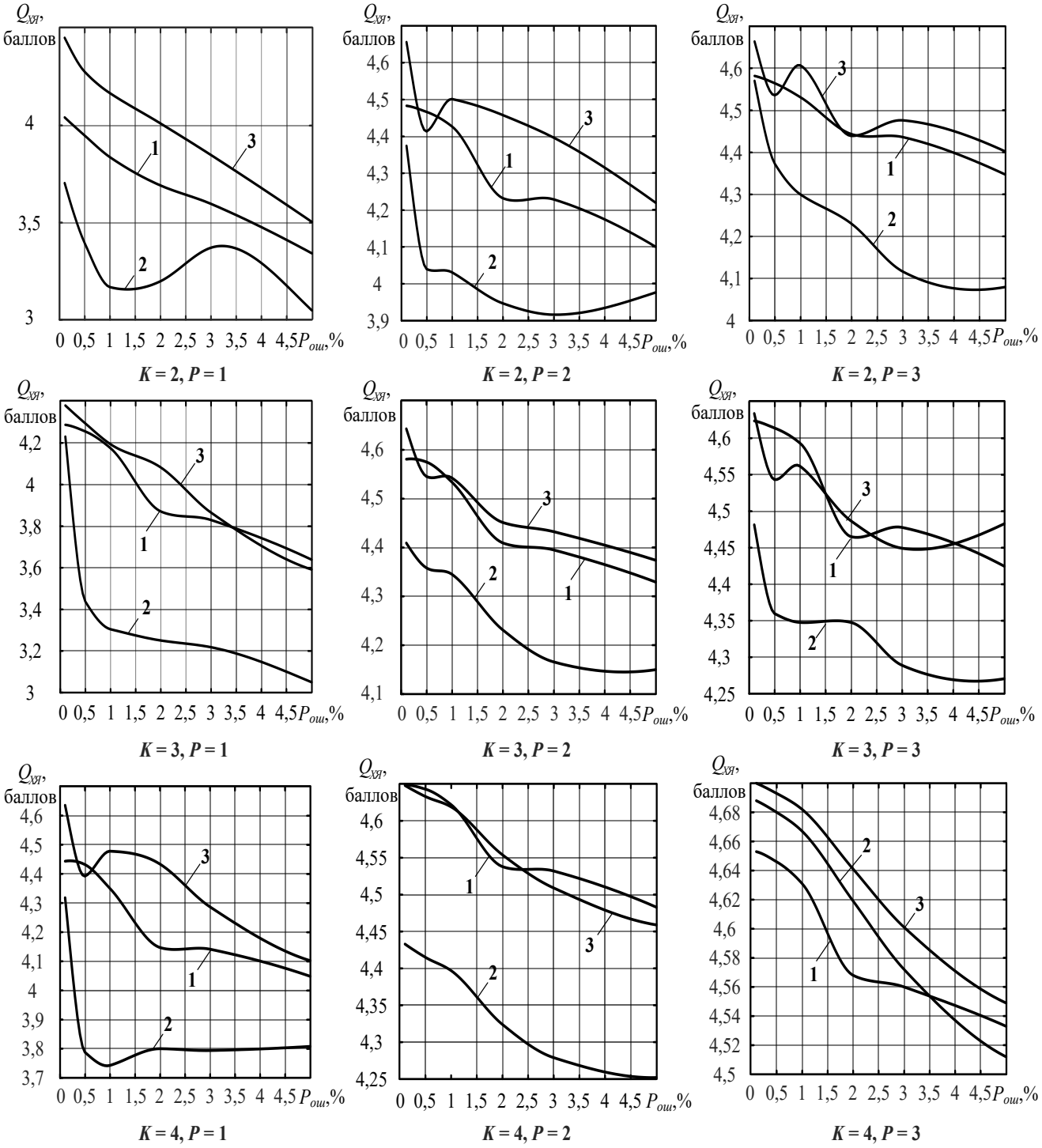


Рисунок 2.13 – Зависимость качества восстановленной речи $Q_{ХЯ}$ от вероятности ошибок в канале $P_{ош}$ для системы на основе представления Хургина – Яковлева при различных конфигурациях (K, P):

1 — MMBE (1,2 кбит/с); 2 — G.723.1 (6,3 кбит/с); 3 — G.726 (24 кбит/с)

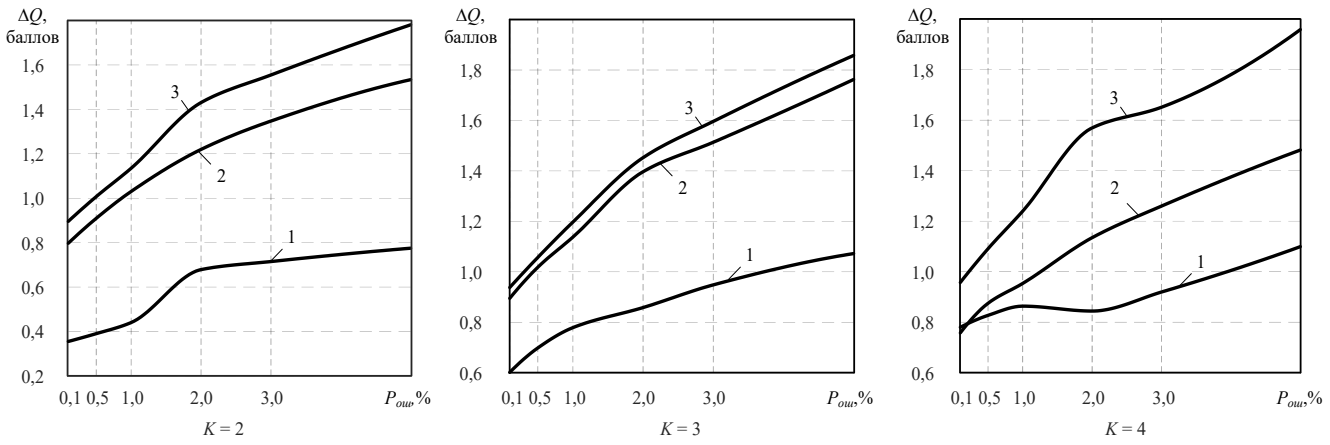
Из представленных зависимостей следует, что с ростом вероятности ошибок $P_{ош}$ качество восстановленной речи уменьшается в обеих системах, однако уменьшение в системе на основе представления Хургина – Яковлева существенно ниже. Это означает, что при высоких $P_{ош}$ расстояние между кривыми, характеризующее выигрыш ΔQ, возрастает, то есть предложенный метод становится тем более эффективным, чем тяжелее условия передачи. Данное свойство принципиально важно для практических применений в системах радиосвязи, где условия передачи изменяются непредсказуемо.

На рисунке 2.14 представлены результаты зависимости выигрыша в помехоустойчивости ΔQ системы кодирования первичных РС на основе предложенного алгоритма с многоэтапным разложением по сравнению с системой, основанной на теореме В.А. Котельникова, при различных вероятностях ошибок на входе первичного декодера.

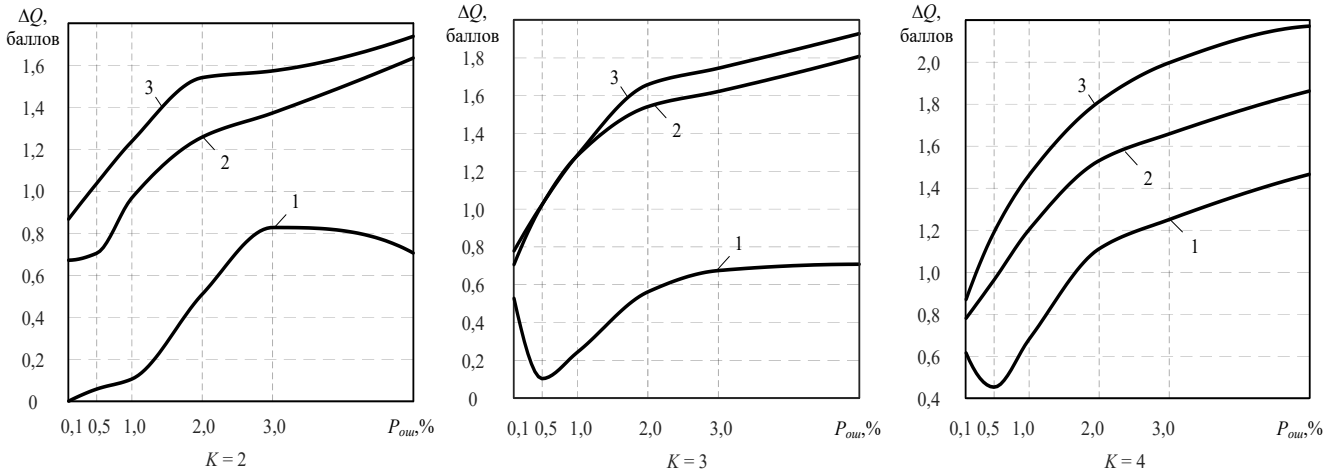
Для анализа использовались три алгоритма первичного кодирования: MMBE со скоростью передачи 1,2 кбит/с (рисунок 2.14а), G.723.1 со скоростью передачи 6,3 кбит/с (рисунок 2.14б) и G.726 со скоростью передачи 24 кбит/с (рисунок 2.14в). На каждом графике цифра 1 соответствует выигрышу в помехоустойчивости системы при одноэтапном разложении (P = 1), цифра 2 — при двухэтапном разложении (P = 2), а цифра 3 — при трехэтапном разложении (P =3).

Анализ результатов экспериментальных зависимостей, приведенных на рисунке 2.14, показывает, что во всех исследованных случаях выигрыш качества восстановленной речи ΔQ остаётся строго положительным, что подтверждает стабильное преимущество предложенного метода при наличии ошибок в канале по сравнению с системой, основанной на теореме В.А. Котельникова.

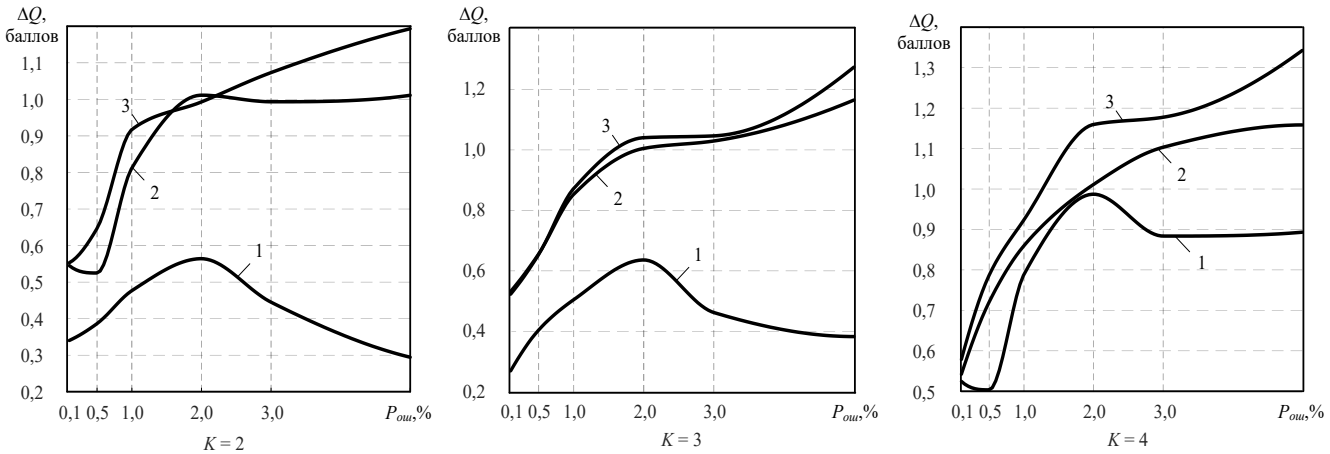
Кроме того, выигрыш возрастает с увеличением вероятности ошибок $P_{ош}$, что особенно важно для применений в условиях интенсивных помех. Это объясняется тем, что чем выше уровень ошибок в канале, тем существеннее вклад механизма фазовой компенсации, реализуемого синтезирующими фильтрами.



а)



б)



в)

Рисунок 2.14 – Выигрыш в качестве восстановленной речи при различных K и P в условиях ошибок канала:
а) ММВЕ (1,2 кбит/с); б) G.723.1 (6,3 кбит/с); в) G.726 (24 кбит/с)]

Для систем на основе кодека ММВЕ (1,2 кбит/с) при наличии ошибок в канале выигрыш в помехоустойчивости изменяется в диапазонах: двухканальная система ($K = 2$, $P = 1…3$) – 0,36…1,78 балла по шкале MOS; трёхканальная система ($K = 3$) – 0,60…1,86 балла; четырёхканальная система ($K = 4$) – 0,79…1,96 балла.

Максимальный выигрыш для данного кодека достигает 1,96 балла по шкале MOS при $K = 4$, $P = 3$ и $P_{ош} = 5\%$.

Для систем на основе кодека G.723.1 (6,3 кбит/с) выигрыш в помехоустойчивости составляет: для двухканальной системы – до 1,74 балла; для трёхканальной – 0,53…1,93 балла; для четырёхканальной – 0,61…2,2 балла. Максимальный выигрыш 2,2 балла (при $K = 4$, $P = 3$) является наибольшим среди всех исследованных кодеков, что объясняется более высокой чувствительностью алгоритма G.723.1 к ошибкам передачи параметров по сравнению с кодеком G.726.

Для систем на основе высокоскоростного кодека G.726 (24 кбит/с) выигрыш в помехоустойчивости составляет: для двухканальной системы – 0,29…1,19 балла; для трёхканальной – 0,27…1,27 балла; для четырёхканальной – 0,50…1,34 балла. Хотя абсолютные значения выигрыша меньше, чем для низкоскоростных кодеков, они остаются существенными для практических применений. Особо следует отметить, что для $K = 2$, $P = 1$ при малых $P_{ош}$ (0,1…0,5 %) выигрыш составляет 0,29 балла, что превышает минимальный порог заметного улучшения качества по шкале MOS.

Сводные результаты экспериментального исследования помехоустойчивости для всех конфигураций и кодеков представлены в таблице 2.6, где приведены диапазоны выигрыша ΔQ (в баллах MOS) при изменении вероятности ошибки от 0,1 до 5 %.

Сравнительный анализ результатов по всем трём кодекам выявляет общую закономерность: наибольший выигрыш в помехоустойчивости наблюдается для низкоскоростного кодека MMBE (1,2 кбит/с) и среднескоростного кодека G.723.1 (6,3 кбит/с). Это объясняется тем, что при сниженных скоростях передачи ошибка в одном бите влечёт за собой искажение целого фрейма длительностью 20…30 мс, тогда как в системе на основе представления Хургина – Яковлева механизм фазовой компенсации позволяет частично нейтрализовать этот эффект. Для высокоскоростного кодека G.726 чувствительность к ошибкам канала ниже в силу методологии кодирования (прямое квантование форм сигнала), тем не менее выигрыш остаётся устойчивым во всём диапазоне исследованных параметров.

Таблица 2.6. Диапазоны выигрыша ΔQ в качестве восстановленной речи при наличии ошибок в канале связи $\left(P_{ош} = 0{,}1...5\%\right)$ для различных конфигураций системы передачи

Кодек	Число каналов K	$P = 1$	$P = 2$	$P = 3$
MMBE (1,2 кбит/с)	$K = 2$	0,36…0,77	0,80…1,53	0,89…1,78
	$K = 3$	0,60…1,07	0,89…1,76	0,94…1,86
	$K = 4$	0,79…1,06	0,77…1,52	0,99…1,96
G.723.1 (6,3 кбит/с)	$K = 2$	0,00…0,65	0,67…1,63	0,87…1,74
	$K = 3$	0,53…0,71	0,71…1,81	0,79…1,93
	$K = 4$	0,61…1,47	0,80…1,85	0,91…2,20
G.726 (24 кбит/с)	$K = 2$	0,29…0,34	0,54…1,01	0,55…1,19
	$K = 3$	0,27…0,64	0,53…1,16	0,52…1,27
	$K = 4$	0,50…0,99	0,54…1,17	0,58…1,34

Обобщение результатов экспериментального исследования, приведённых в таблице 2.6 и на рисунках 2.14, позволяет сформулировать следующее. Алгоритмы первичного кодирования РС, реализованные на основе представления Хургина – Яковлева с многоэтапным разложением, обеспечивают повышение качества восстановленной речи по шкале MOS до 2,2 балла для кодека G.723.1 (6,3 кбит/с) и до 1,96 балла для кодека MMBE (1,2 кбит/с) при вероятности битовых ошибок в канале до 5 %. Выигрыш достигается без изменения скоростей передачи соответствующих кодеков, поскольку суммарный объём передаваемой информации в системе на основе представления Хургина – Яковлева остаётся неизменным: K-кратное снижение числа отсчётов сигнала компенсируется передачей $\left(K-1\right)$ производных. Таким образом, экспериментально подтверждается один из основных научных результатов настоящей диссертации: алгоритмы первичного кодирования РС на основе представления Хургина – Яковлева обеспечивают повышение качества восстановленной речи по шкале MOS на величину до 2,2 балла в системах передачи речи различной скорости при вероятности битовых ошибок в канале до 5 % без снижения скорости передачи.

2.5 Выводы

В второй главе разработаны теоретические и алгоритмические основы реализации системы первичного кодирования РС на основе представления Хургина – Яковлева применительно к радиотехническим системам передачи информации, проведён анализ эквивалентности с банками фильтров, исследована помехоустойчивость при воздействии аддитивного белого гауссовского шума и проведены экспериментальные исследования эффективности предложенного метода совместно с алгоритмами первичного кодирования различных классов скоростей передачи. По результатам выполненных исследований сформулированы следующие выводы.

1. Установлена эквивалентность системы обработки и передачи РС на основе представления Хургина – Яковлева банку фильтров с децимацией и интерполяцией в K раз. Показано, что передаточные функции каналов анализа и синтеза в предложенной системе допускают полифазное разложение, позволяющее сократить вычислительную сложность пропорционально коэффициенту децимации.

2. Проведён сравнительный анализ вычислительной сложности предложенной системы и классической системы дискретизации на основе теоремы В.А. Котельникова. Установлено, что реализация многоканальной и многоуровневой структуры на основе полифазного разложения позволяет снизить вычислительные затраты по сравнению с непосредственной реализацией синтезирующих фильтров классического алгоритма. Для двухканальной системы ($K = 2$) при числе этапов разложения $P = 1…3$ относительное сокращение числа операций умножения достигает диапазона от 1,143 до 1,75 раза для кадров анализа размером $M = 128$ и $M = 256$ соответственно. С увеличением числа каналов K и числа этапов P выигрыш в вычислительной сложности дополнительно возрастает, достигая для конфигурации $K = 4$, $P = 3$ значения до 7 раз при $M = 128$.

3. Аналитически обоснован выигрыш в помехоустойчивости системы при воздействии аддитивного белого гауссовского шума. Показано, что дополнительные фазовые сдвиги, вносимые синтезирующими фильтрами системы, обеспечивают геометрическое суммирование составляющих шума в каналах с

различными фазовыми соотношениями, что приводит к частичной взаимной компенсации шумовых составляющих на выходе суммирующего устройства. Расчётный выигрыш в ОСШ составляет: для одноэтапного разложения — 2,27 дБ (K = 2), 6,2 дБ (K = 3) и 8,68 дБ (K = 4); при многоэтапном разложении выигрыш возрастает пропорционально числу этапов P.

4. Экспериментально установлено, что применение алгоритмов первичного кодирования РС на основе представления Хургина – Яковлева с многоэтапным разложением обеспечивает устойчивый выигрыш в качестве восстановленной речи по шкале MOS во всём исследованном пространстве параметров и при всех рассмотренных алгоритмах первичного кодирования, как в условиях идеального канала, так и при наличии ошибок передачи. Увеличение числа каналов K и числа этапов разложения P увеличивает помехоустойчивость системы и качество восстановленной речи на приёме. Максимальный выигрыш достигается при использовании кодека G.723.1 (6,3 кбит/с) и конфигурации $K = 4$, $P = 3$ при $P_{ош} = 5\%$ и составляет 2,2 балла по шкале MOS.

ГЛАВА 3 ИССЛЕДОВАНИЕ ИНТЕГРИРОВАННОЙ СИСТЕМЫ КОДИРОВАНИЯ РЕЧИ НА ОСНОВЕ ПРЕДСТАВЛЕНИЯ ХУРГИНА-ЯКОВЛЕВА И ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ

3.1 Вводные замечания

Задача низкоскоростной передачи речи при сохранении приемлемого качества воспроизведения относится к числу фундаментальных задач в радиотехнике. Диапазон скоростей 1…5 кбит/с, характерный для тактических, мобильных и аварийно-спасательных систем связи, предъявляет к алгоритмам обработки требования, которые не могут быть удовлетворены методами первичного кодирования, основанными исключительно на снижении частоты дискретизации. При любой разумной конфигурации многоканальной системы обработки поток передаваемых отсчётов остаётся существенно выше целевой скорости кодека, и, следовательно, необходим самостоятельный этап сжатия, способный радикально разредить этот поток без деградации существенного уменьшения качества речи.

ДВП является одним из наиболее теоретически обоснованных инструментов для решения данной задачи. В отличие от преобразования Фурье, оно обеспечивает одновременное разрешение по времени и частоте, что принципиально важно для нестационарных РС с быстро меняющимися переходными участками [86, 87]. Многоуровневое вейвлет-разложение концентрирует основную энергию речевого сигнала в сравнительно малом числе значимых коэффициентов, тогда как значения большинства коэффициентов оказывается пренебрежимо малыми. Данное свойство естественной разреженности создаёт непосредственную основу для порогового сжатия: коэффициенты, энергия которых не превышает некоторого порога, могут быть обнулены без ощутимых потерь качества восстановленного сигнала [88].

Практическая эффективность порогового сжатия критически зависит от способа выбора самих порогов. Классические методы — универсальный порог Донохо–Джонстона и критерий нерискованной оценки по Штейну (SUREShrink) — получены в предположении о статистической однородности шума и приближённой гауссовости вейвлет-коэффициентов [51–53]. РС не удовлетворяют этим предположениям: их энергетическое распределение по уровням разложения

существенно неравномерно и изменяется от кадра к кадру. Применение фиксированных порогов либо влечёт чрезмерное обнуление коэффициентов с деградацией разборчивости, либо обеспечивает недостаточную разреженность. Следовательно, задача оптимального выбора порогов применительно к речевым сигналам не может быть решена в рамках существующих универсальных подходов и требует разработки специализированного метода, учитывающего энергетическую структуру речи на каждом уровне разложения.

Для решения этой задачи в настоящей работе предлагается метод оптимизации энергетических порогов на основе алгоритма роя частиц PSO (Particle Swarm Optimization). Выбор данного алгоритма обусловлен рядом принципиальных соображений. Целевая функция задачи оптимизации порогов является негладкой и многоэкстремальной вследствие кусочно-постоянного характера операции обнуления коэффициентов, что исключает применение градиентных методов. Алгоритм роя частиц PSO принадлежит к классу метаэвристических алгоритмов глобальной оптимизации, не требующих вычисления градиента, допускающих многокритериальные ограничения непосредственно в структуре целевой функции и демонстрирующих устойчивую сходимость в пространствах умеренной размерности [57–70]. При числе уровней разложения $J = 4$ размерность пространства поиска составляет $J + 1 = 5$ для одноканальной системы и $2(J + 1) = 10$ для двухканальной, что соответствует диапазону эффективного применения данного алгоритма.

Совместное применение представления Хургина – Яковлева и ДВП в едином тракте обработки образует интегрированную систему, в которой каждый из двух компонентов устраняет ограничения другого. Представление Хургина – Яковлева снижает требуемую скорость дискретизации и повышает помехоустойчивость за счёт фазовых характеристик синтезирующих фильтров; ДВП с оптимальными энергетическими порогами обеспечивает максимальную разреженность представления при заданных ограничениях на качество восстановленной речи. Совместная оптимизация порогов обоих каналов методом роя частиц позволяет полностью реализовать потенциал такой архитектуры,

который не достигается при независимой настройке каждого из компонентов.

Вместе с тем, практическая ценность предложенных алгоритмов определяется не только их теоретической эффективностью, но и реализуемостью в условиях ограниченных аппаратных ресурсов встраиваемых радиотехнических систем. Анализ вычислительной сложности интегрированной системы и обоснование рекомендаций по её реализации на современной элементной базе являются необходимым условием перехода от исследовательского прототипа к практически реализуемому решению.

Предметом настоящей главы является разработка и исследование метода сжатия РС на основе ДВП с оптимизацией энергетических порогов алгоритмом роя частиц, исследование интегрированной системы совместной обработки и анализ практических аспектов её реализации на современной элементной базе.

В третьей главе диссертационной работы решаются следующие задачи:

1. Разработка метода сжатия РС на основе покадрового ДВП с энергетическим критерием порогового сжатия, включая обоснованный выбор оптимального вейвлета и числа уровней разложения.

2. Постановка задачи многокритериальной оптимизации вектора энергетических порогов и разработка алгоритма её решения на основе метода роя частиц PSO; верификация метода в сравнении с классическими методами порогового сжатия.

3. Исследование интегрированной системы на основе представления Хургина – Яковлева и ДВП, включая совместную оптимизацию параметров двух каналов и сравнительный анализ с низкоскоростными кодеками при наличии ошибок в канале связи.

4. Анализ вычислительной сложности алгоритмов интегрированной системы и формулирование рекомендаций по реализации на современных цифровых сигнальных процессорах и программируемых логических интегральных схемах.

3.2 Метод сжатия речевых сигналов на основе покадрового дискретного вейвлет-преобразования

3.2.1 Покадровое дискретное вейвлет-преобразование речевых сигналов

РС является нестационарным случайным процессом, локальные статистические характеристики которого значительно изменяются на интервалах порядка 10…40 мс. Применение единого глобального преобразования ко всей записи неизбежно усредняет эту нестационарность и снижает эффективность последующего сжатия. Для устранения данного ограничения вводится покадровая схема обработки.

Исходный дискретный сигнал $x[n]$, где $n = 0,1,\ldots,N-1,$ длиной N отсчётов, разбивается на L перекрывающихся кадров длиной $M = 256$ отсчётов (что соответствует 32 мс при частоте дискретизации 8 кГц). Каждый кадр $y_m[n]$, где $m = 1,2,\ldots,L,$ взвешивается окном Ханна:

$$y_m[n] = x\left[n + (m-1)S\right]w[n],\ w[n] = \sin^2\left(\frac{\pi n}{M}\right),\ n = 0,1,\ldots,M-1, \tag{3.1}$$

где S – шаг сдвига кадра. Применение окна Ханна подавляет краевые разрывы и спектральные утечки, неизбежные при работе с конечными фрагментами нестационарного сигнала, и тем самым повышает точность вейвлет-разложения на переходных участках речи.

К каждому кадру независимо применяется J-уровневое ДВП по быстрому алгоритму Маллата [30]. На каждом уровне $j = 1,2,\ldots,J$ входная последовательность фильтруется парой сопряжённых квадратурных фильтров – низкочастотным $h[n]$ и высокочастотным $g[n]$ – с последующей децимацией с коэффициентом 2:

$$cA_j^{(m)}[k] = \sum_{n\in\mathbb{Z}} h[n-2k]\,cA_{j-1}^{(m)}[n] = \left(h * cA_{j-1}^{(m)}\right)\downarrow 2, \tag{3.2}$$

$$cD_j^{(m)}[k] = \sum_{n\in\mathbb{Z}} g[n-2k]\,cA_{j-1}^{(m)}[n] = \left(g * cA_{j-1}^{(m)}\right)\downarrow 2, \tag{3.3}$$

где $cA_j^{(m)}[k]$ – коэффициенты аппроксимации уровня j кадра m (низкочастотная составляющая), $cD_j^{(m)}[k]$ – коэффициенты деталей уровня j кадра m (высокочастотная составляющая), $h[n]$ и $g[n]$ – импульсные характеристики низкочастотного и высокочастотного фильтров анализа, связанные с выбранным

вейвлетом, символ $(*)$ обозначает операцию свертки, $\downarrow 2$ – децимацию с коэффициентом 2 (сохранение четных отсчетов). Начальное условие: $cA_0^{(m)}[n] = y_m[n]$.

Итеративное применение (3.2), (3.3) для $j = 1,2,\ldots,J$ формирует полное J-уровневое разложение кадра $y_m[n]$. На каждом уровне коэффициенты аппроксимации $cA_j^{(m)}$ подвергаются дальнейшей декомпозиции, в то время как коэффициенты деталей $cD_j^{(m)}$ сохраняются. Результатом является набор вейвлет-коэффициентов кадра m $\left\{cA_J^{(m)}, cD_J^{(m)}, cD_{J-1}^{(m)}, \ldots, cD_1^{(m)}\right\}$, где $cA_J^{(m)}$ представляет низкочастотную аппроксимацию на самом грубом масштабе, а $cD_j^{(m)}$ – детали на последовательно убывающих масштабах (возрастающих частотах). Схема многоуровневого ДВП кадра приведена на рисунке 3.1.

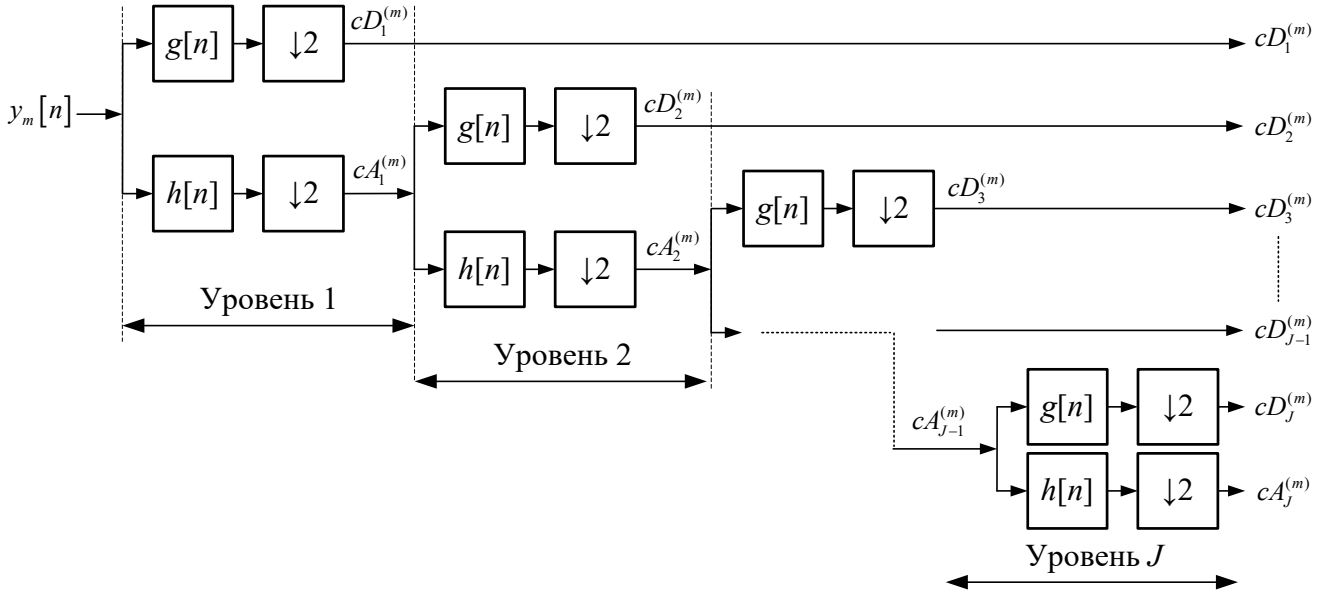


Рисунок 3.1 – Схема многоуровневого ДВП кадра

По завершении J итераций формируется полный вектор вейвлет-коэффициентов кадра m:

$$\mathbf{C}_m = \left[cA_J^{(m)}; cD_J^{(m)}; cD_{J-1}^{(m)}; \ldots; cD_1^{(m)} \right]^T, \tag{3.4}$$

имеющий длину M отсчётов при использовании периодических граничных условий. Параметр J определяет глубину частотного разрешения: каждый дополнительный уровень разложения вдвое уменьшает частотную полосу

аппроксимирующих коэффициентов, концентрируя в них всё большую долю энергии речевого сигнала.

Ключевым свойством такого представления является неравномерное распределение энергии по уровням разложения. Для коэффициентов уровня *l* кадра *m* определим полную энергию уровня:

$$E_l^{(m)} = \sum_{k=1}^{L_l} \left(c_k^{(l,m)} \right)^2, \tag{3.5}$$

где L_l – число коэффициентов на уровне *l* кадра *m;* $c_k^{(l,m)}$ – *k*-й коэффициент уровня *l* кадра *m* $\left(k = 1,2,\ldots,L_l\right)$. Для РС низкочастотные уровни (*l* малые, аппроксимирующие) аккумулируют до 90…95 % суммарной энергии, тогда как высокочастотные уровни деталей содержат преимущественно малозначимые коэффициенты. Это обстоятельство составляет физическую основу порогового сжатия.

Восстановление сигнала для каждого кадра осуществляется посредством обратного ДВП с использованием дуальных синтезирующих фильтров $\tilde{h}[n]$ и $\tilde{g}[n]$:

$$\hat{y}_m[n] = \mathrm{IDWT}_J\left(\tilde{\mathbf{C}}_m\right), \tag{3.6}$$

где $\tilde{\mathbf{C}}_m$ — вектор коэффициентов после порогового сжатия. Итоговый восстановленный сигнал $\hat{x}[n]$ формируется путём взвешенного перекрывающего-складывающего суммирования обработанных кадров.

3.2.2 Классические методы пороговой обработки вейвлет-коэффициентов

Задача выбора порогов для обнуления незначимых вейвлет-коэффициентов исследовалась в литературе преимущественно в контексте удаления шума, а не сжатия. Тем не менее именно эти методы образуют естественную базу сравнения для оценки эффективности предложенного подхода. В данном разделе систематизированы основные классические методы порогового сжатия применительно к задаче кодирования РС; анализируются их допущения и ограничения.

Независимо от конкретного критерия выбора порога, все классические методы используют одну из двух стандартных функций порогового отображения [51]:

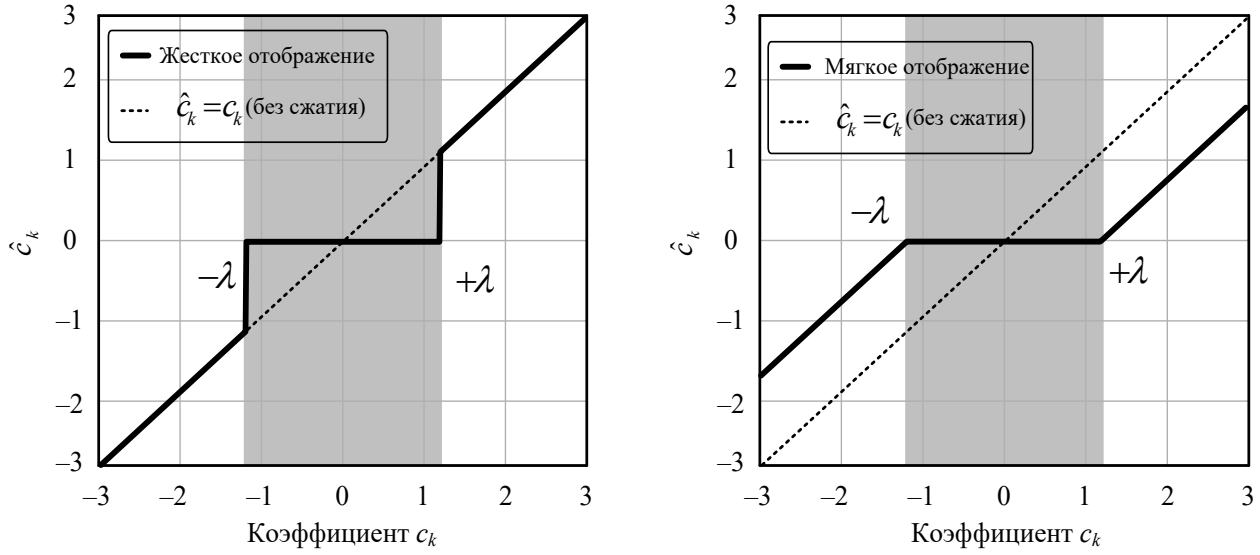
– жёсткое пороговое отображение:

$$\hat{c}_k = \begin{cases} c_k, & |c_k| > \lambda, \\ 0, & |c_k| \le \lambda. \end{cases} \tag{3.7}$$

– мягкое пороговое отображение:

$$\hat{c}_k = \begin{cases} \operatorname{sgn}(c_k)\left(|c_k| - \lambda\right), & |c_k| > \lambda, \\ 0, & |c_k| \le \lambda, \end{cases} \tag{3.8}$$

где $\lambda > 0$ – амплитудный порог; c_k и $\hat{c}_k$ – исходный и обработанный коэффициенты соответственно.



а) б)

Рисунок 3.2 – Функции жёсткого (а) и мягкого (б) порогового отображения

Из рисунка 3.2 следует принципиальное различие двух отображений. Жёсткое (3.7) сохраняет точные значения значимых коэффициентов, однако порождает разрывы в точках $\pm\lambda$, проявляющиеся как высокочастотные артефакты в восстановленном сигнале. Мягкое (3.8) устраняет разрывы, но вносит систематическое смещение $-\lambda$ для всех коэффициентов $|c_k| > \lambda$, что эквивалентно занижению амплитуд значимых компонент речи. Принципиальное ограничение обоих отображений состоит в том, что порог λ применяется равномерно ко всем уровням разложения без учёта частотно-зависимого энергетического распределения речевого сигнала.

Рассмотрим первый метод. Универсальный порог Донохо–Джонстона (VisuShrink). Донохо и Джонстон [52] предложили выбирать единый порог для всех

уровней разложения, исходя из следующего соображения: коэффициенты с малой амплитудой, подлежащие обнулению, могут быть смоделированы как гауссовский белый шум с дисперсией σ^2. При этом допущении порог можно записать в следующем виде:

$$\lambda^* = \hat{\sigma}\sqrt{2\ln C_{tot}}. \tag{3.9}$$

Он является асимптотически оптимальным в минимаксном смысле. Здесь C_{tot} – общее число коэффициентов, $\hat{\sigma}$ – оценка среднеквадратического отклонения, вычисляемая по медианному оценщику на первом уровне деталей:

$$\hat{\sigma} = \frac{\text{median}\left(\left|c_k^{(l,m)}\right|\right)}{0,6745}, \tag{3.10}$$

где $c_k^{(l,m)}$ – коэффициенты детализации первого уровня m-го кадра. Порог (3.9) не зависит от уровня разложения и применяется равномерно ко всем коэффициентам.

Данный метод имеет два теоретических ограничения. Во-первых, гауссова модель малоамплитудных коэффициентов точна лишь в центральной области распределения; для РС характерно распределение с тяжёлыми хвостами [53], что приводит к систематической переоценке $\hat{\sigma}$ при применении оценщика по формуле (3.10) к уровням с иным характером распределения. Во-вторых, единственность порога λ^* означает, что он не адаптируется к различному энергетическому содержанию уровней разложения: коэффициенты нижних уровней, несущих основную энергию речевого сигнала, и коэффициенты верхних уровней с пренебрежимой энергией обрабатываются по одному критерию, что заведомо не является оптимальным.

Рассмотрим второй метод на основе адаптивного порога SURE Shrink. Для устранения зависимости от единого порога Донохо и Джонстон [52] предложили выбирать порог адаптивно для каждого уровня разложения, минимизируя несмещённую оценку риска Штейна (Stein's Unbiased Risk Estimate — SURE):

$$\lambda_l^{SURE} = \arg\ \min_{\lambda \geq 0} \text{SURE}\left(\lambda, c^{(l)}\right), \tag{3.11}$$

где оценка среднеквадратической ошибки реконструкции при пороге λ на уровне l имеет следующий вид:

$$\mathrm{SURE}\left(\lambda, \mathbf{c}^{(l)}\right) = L_l - 2\sum_{k=1}^{L_l} \mathbf{1}\left[\left|c_k^{(l)}\right| \le \lambda\right] + \sum_{k=1}^{L_l} \min\left(\left(c_k^{(l)}\right)^2, \lambda^2\right), \tag{3.12}$$

где L_l – число коэффициентов уровня l; $\mathbf{1}[\cdot]$ – функция-индикатор, равная единице при выполнении условия. В отличие от VisuShrink, SUREShrink адаптирует порог к каждому уровню, что теоретически позволяет точнее учесть различие в распределении коэффициентов. Вместе с тем, принципиальное ограничение метода состоит в несоответствии целевого критерия: минимизация СКО реконструкции и максимизация разреженности ρ при ограничениях на качество — это принципиально различные задачи оптимизации. Минимизация СКО побуждает алгоритм к консервативному выбору порога, сохраняющему избыточное число коэффициентов, тогда как задача сжатия требует находить наиболее агрессивный порог, при котором показатели качества ещё остаются в допустимой области.

Третий метод построен на основе фиксированного энергетического порога. В отличие от двух предыдущих подходов, применение единого порогового коэффициента $\alpha_l = \alpha_0$ для всех уровней разложения переводит задачу отбора коэффициентов в энергетическую плоскость: на каждом уровне сохраняется фиксированная доля α_0 его энергии независимо от амплитудного распределения коэффициентов. Такой подход более адекватен для РС, поскольку критерий качества восстановления речи (SegSNR, ViSQOL) определяется энергетикой ошибки, а не поточечными отклонениями амплитуд. Вместе с тем единственность α_0 для всех уровней не учитывает их различную информативность: оптимальная доля сохраняемой энергии объективно должна быть выше на нижних уровнях, несущих просодическую и форматную структуру речи и ниже на верхних уровнях с пренебрежимым перцептуальным вкладом. Применение единого α_0 означает, что либо нижние уровни сжимаются недостаточно агрессивно, либо верхние сжимаются недостаточно — в зависимости от выбранного значения α_0.

Рассмотренные методы обнаруживают последовательную логику развития: от единого амплитудного порога (VisuShrink) к уровнево-адаптивному (SUREShrink) и далее к энергетическому критерию (фиксированный α_0). Каждый

шаг устраняет одно ограничение предшественника, однако вводит новое. Ни один из методов не обеспечивает одновременно три свойства, необходимые для эффективного сжатия РС: уровневую адаптацию порогов, энергетический критерий их выбора и прямую оптимизацию разреженности при явных ограничениях на качество восстановления. Разработка метода, сочетающего все три свойства, показана в разделе 3.3.

3.2.3 Алгоритм энергетического порогового сжатия вейвлет-коэффициентов

Общий принцип порогового сжатия состоит в обнулении тех коэффициентов, которые признаются малозначимыми по некоторому критерию. Выбор этого критерия принципиален. Классические методы — универсальный порог Донохо–Джонстона и SUREShrink [51, 52] — оценивают значимость каждого коэффициента по его абсолютной величине в предположении о гауссовости шума и статистической однородности коэффициентов. РС данному предположению не удовлетворяют: их коэффициенты на разных уровнях разложения имеют принципиально разные дисперсии и вклад в суммарную энергию. В результате применение амплитудного порога к уровням с существенно разными энергиями неизбежно приводит либо к избыточному сжатию низкочастотных уровней, либо к недостаточному — высокочастотных.

В предлагаемом методе используется энергетический критерий: на каждом уровне l сохраняется не фиксированное число коэффициентов по амплитуде, а та минимальная их совокупность, которая обеспечивает сохранение заданной доли $\alpha_l \in [0,1]$ полной энергии уровня. Такой подход автоматически адаптируется к энергетической структуре уровня: на насыщенных энергией уровнях сохраняется больше коэффициентов, на малозначимых — меньше. Алгоритм применяется независимо к каждому из $J+1$ уровней разложения (аппроксимация при $l=0$ и уровни деталей при $l=1,\ldots,J$) каждого кадра m и состоит из четырёх шагов.

Шаг 1. Коэффициенты уровня l ранжируются по убыванию абсолютных значений с формированием перестановки $\Pi(\cdot)$:

$$\left|c_{\Pi(1)}^{(l,m)}\right| \geq \left|c_{\Pi(2)}^{(l,m)}\right| \geq \ldots \geq \left|c_{\Pi(L_l)}^{(l,m)}\right|. \tag{3.13}$$

Шаг 2. Вычисляется кумулятивная энергия первых s коэффициентов упорядоченного набора:

$$E_{\Sigma}^{(l,m)}(s) = \sum_{\xi=1}^{s}\left(c_{\Pi(\xi)}^{(l,m)}\right)^2, \tag{3.14}$$

где ξ – индекс суммирования, используемый локально только в данной формуле $\left(\xi = 1,\ldots,s\right)$.

Шаг 3. Определяется минимальный индекс s^*, при котором накопленная энергия достигает заданной доли α_l от полной энергии уровня:

$$s^* = \min\left\{ s \in \{1,\ldots,L_l\} : \frac{E_{\Sigma}^{(l,m)}(s^*)}{E_l^{(m)}} \geq \alpha_l \right\}. \tag{3.15}$$

Шаг 4. Формируется вектор сжатых коэффициентов: сохраняются s^* наиболее энергозначимых коэффициентов, остальные обнуляются:

$$\tilde{c}_k^{(l,m)} = \begin{cases} c_k^{(l,m)}, \text{ если } \left|c_k^{(l,m)}\right| \geq \left|c_{\Pi(s^*)}^{(l,m)}\right|, \\ 0, \text{ иначе.} \end{cases} \tag{3.16}$$

Стратегия сжатия целиком определяется вектором энергетических порогов:

$$\boldsymbol{\alpha} = \left[\alpha_0, \alpha_1, \ldots, \alpha_J\right]^T \in \left[0,1\right]^{J+1}. \tag{3.17}$$

Физический смысл параметра α_l однозначен: он задаёт целевую долю энергии уровня l, подлежащую сохранению после сжатия. При $\alpha_l \to 1$ все коэффициенты уровня сохраняются; при $\alpha_l \to 0$ уровень полностью обнуляется. Таким образом, вектор энергетических порогов $\boldsymbol{\alpha}$ обеспечивает раздельное управление степенью сжатия в каждой частотной полосе, что принципиально отличает предлагаемый метод от подходов с единым глобальным порогом.

Процедура покадровой обработки сигнала с энергетическим пороговым сжатием, описанная выше, показана на рисунке 3.3. Схема иллюстрирует полный цикл: от разложения сигнала до восстановления, с указанием вектора порогов $\boldsymbol{\alpha}$, как ключевого элемента управления степенью сжатия.

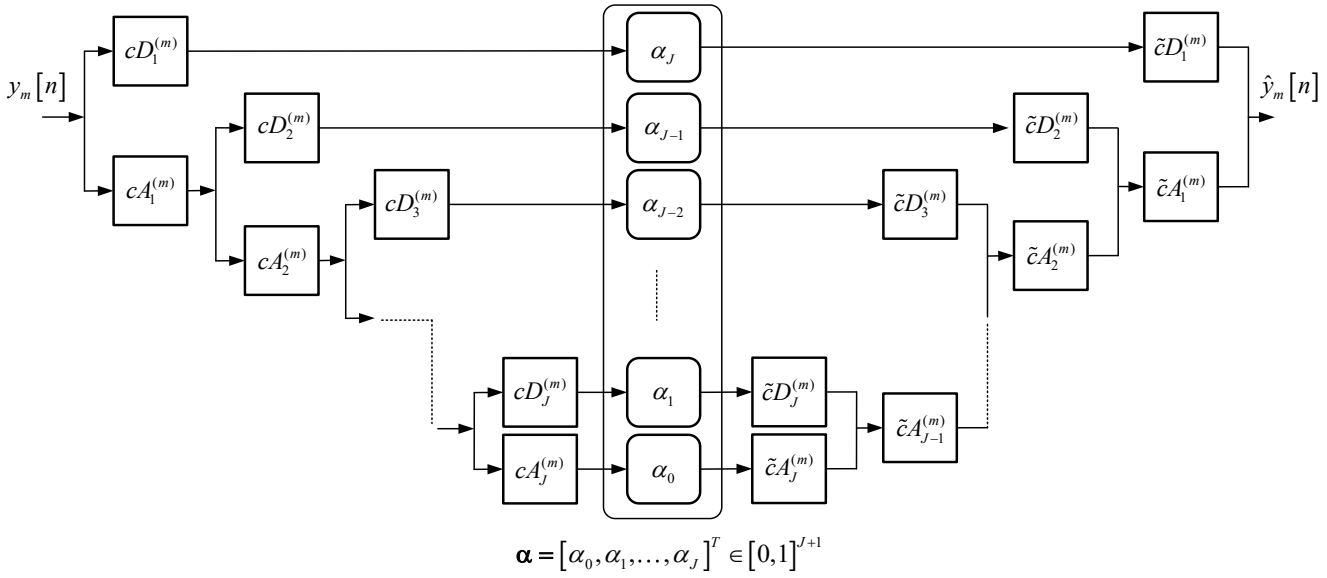


Рисунок 3.3 – Схема энергетического порогового сжатия сигнала на основе ДВП

3.2.4 Метрики эффективности сжатия

Выбор метрик определяется природой задачи: сжатие создаёт компромисс между разреженностью представления и точностью восстановления сигнала, причём точность должна оцениваться как в объективном акустическом, так и в перцептуальном смысле. Для полного охвата этого компромисса используется система из трёх метрик.

– Коэффициент сжатия ρ количественно характеризует степень сжатия как долю обнулённых коэффициентов по всем кадрам и уровням:

$$\rho(\boldsymbol{\alpha}) = \left(1 - \frac{C_{nz}(\boldsymbol{\alpha})}{C_{tot}}\right).100\%, \tag{3.18}$$

где $C_{tot} = ML$ – общее число вейвлет-коэффициентов всех M кадров; $C_{nz}(\boldsymbol{\alpha}) = \sum_{m=1}^{M}\sum_{l=0}^{J} s_l^{*(m)}(\boldsymbol{\alpha})$ – общее число ненулевых коэффициентов после порогового сжатия. Высокие значения коэффициент сжатия $\rho(\boldsymbol{\alpha})$, близкие к 100 %, свидетельствуют об эффективном сжатии сигнала.

– Сегментное отношение сигнал–шум Q_1 является основной объективной акустической метрикой. В отличие от глобального SNR, рассчитанного по всей записи, Q_1 вычисляется раздельно по R непересекающимся сегментам длиной L_{seg} отсчётов:

$$\mathrm{SNR}_r = 10\log_{10}\left(\frac{\sum_{n\in S_r} x^2[n]}{\sum_{n\in S_r}\left(x[n]-\hat{x}[n]\right)^2}\right)\ [\text{дБ}], \tag{3.19}$$

$$Q_1(\boldsymbol{\alpha}) = \frac{1}{R}\sum_{r=1}^{R} SNR_r\ \ [\text{дБ}]. \tag{3.20}$$

где S_r – множество индексов отсчётов *r*-го сегмента. Сегменты с энергией сигнала менее 10^{-10} исключаются из усреднения, либо им присваивается значение $\mathrm{SNR}_r = 60[\text{дБ}]$. Сегментирование принципиально для оценки качества речи: агрессивное пороговое сжатие вызывает локальные артефакты на переходных участках, которые глобальный SNR маскирует, но Q_1 фиксирует.

– Метрика ViSQOL Q_2 моделирует субъективное восприятие качества речи слушателем на основе психоакустической модели [19]:

$$Q_2(\boldsymbol{\alpha}) = \ \mathrm{ViSQOL}(x[n], \hat{x}[n]), \tag{3.21}$$

где выходное значение Q_2 соответствует шкале среднего мнения MOS в диапазоне 1…5 баллов. Необходимость включения Q_2 наряду Q_1 обусловлена тем, что при нестационарных искажениях оба показателя несут неперекрывающуюся информацию: высокое Q_1 не гарантирует высокого Q_2 при специфических артефактах порогового сжатия, ощутимых слухом, но слабо влияющих на среднюю энергию погрешности. Совместное использование трёх метрик позволяет формализовать задачу управления вектором **α**, как задачу оптимизации: максимизировать разреженность ρ при соблюдении ограничений на акустическое и перцептуальное качество восстановленного сигнала:

$$\rho(\boldsymbol{\alpha}) \to \max_{\boldsymbol{\alpha}\in[0,1]^{J+1}}, \tag{3.22}$$

при ограничениях: $Q_1(\boldsymbol{\alpha}) \ge Q_1^{\min} = 9{,}5$ дБ, $Q_2(\boldsymbol{\alpha}) \ge Q_2^{\min} = 3{,}5$ балла. (3.23)

Пороговые значения $Q_1^{\min}$ и $Q_2^{\min}$ установлены в соответствии с рекомендацией МСЭ-Т P.800 [72] для речи телефонного качества. Задача (3.22), (3.23) является задачей нелинейной оптимизации с негладкой многоэкстремальной

целевой функцией: операция определения s^* по формуле (3.15) является кусочно-постоянной по α_l, что исключает применение градиентных методов. Разработке метода решения этой задачи посвящён следующий раздел.

3.3 Оптимизация энергетических порогов вейвлет-преобразования алгоритмом роя частиц

3.3.1 Постановка задачи многокритериальной оптимизации

Задача, сформулированная в разделе 3.2, состоит в нахождении вектора $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \ldots, \alpha_J]^T \in [0,1]^{J+1}$, максимизирующего коэффициент разреженности $\rho(\boldsymbol{\alpha})$ при выполнении ограничений на качество восстановленной речи. В явном виде она записывается как:

$$\boldsymbol{\alpha}^{\text{opt}} = \max_{\boldsymbol{\alpha} \in [0,1]^{J+1}} F(\boldsymbol{\alpha}), \tag{3.24}$$

где целевая функция $F(\boldsymbol{\alpha})$ представляет собой взвешенную сумму нормализованных показателей эффективности:

$$F(\boldsymbol{\alpha}) = \mathrm{w}_1 f_\rho(\boldsymbol{\alpha}) + \mathrm{w}_2 f_{Q_1}(\boldsymbol{\alpha}) + \mathrm{w}_3 f_{Q_2}(\boldsymbol{\alpha}), \tag{3.25}$$

где: $f_\rho(\boldsymbol{\alpha}) = \rho(\boldsymbol{\alpha})$ – коэффициент сжатия, определённый в (3.18), выраженный в процентах; $f_{Q_1}(\boldsymbol{\alpha}), f_{Q_2}(\boldsymbol{\alpha})$ – нормализованные метрики качества SegSNR и ViSQOL соответственно; а весовые коэффициенты $\mathrm{w}_1, \mathrm{w}_2, \mathrm{w}_3 \geq 0$ удовлетворяют условию $\mathrm{w}_1 + \mathrm{w}_2 + \mathrm{w}_3 = 1$.

При этом должны выполняться ограничения: $Q_1(\boldsymbol{\alpha}) \geq Q_1^{\min}, Q_2(\boldsymbol{\alpha}) \geq Q_2^{\min}$, где $Q_1(\boldsymbol{\alpha})$ и $Q_2(\boldsymbol{\alpha})$ — метрики качества SegSNR и ViSQOL соответственно.

Задача (3.24) принадлежит к классу задач нелинейного программирования с негладкой многоэкстремальной целевой функцией. Негладкость обусловлена кусочно-постоянным характером операции определения порогового индекса s^* в (3.15): при изменении любой компоненты α_l непрерывно целевая функция $F(\boldsymbol{\alpha})$ меняется скачкообразно в точках, где s^* переходит к следующему значению. Это свойство исключает применение градиентных методов оптимизации. Многоэкстремальность обусловлена взаимодействием ограничений: зависимости

$Q_1(\boldsymbol{\alpha})$ и $Q_2(\boldsymbol{\alpha})$ от вектора порогов нелинейны и не выпуклы, что порождает множество локальных оптимумов в пространстве поиска $[0,1]^{J+1}$.

Изложенные свойства делают обоснованным применение метаэвристических алгоритмов глобальной оптимизации. В настоящей работе используется алгоритм роя частиц PSO [57], обладающий рядом преимуществ применительно к данной задаче: не требует вычисления градиента целевой функции; допускает естественное включение ограничений (3.17) непосредственно в структуру целевой функции; демонстрирует устойчивую сходимость в многомерных пространствах умеренной размерности; обладает простой и параллелизуемой вычислительной структурой.

3.3.2 Алгоритм роя частиц: структура, параметры и критерии останова

Алгоритм роя частиц моделирует коллективный поиск оптимального решения группой агентов — частиц — в пространстве параметров [57, 58]. Каждая частица i $(i = 1,2,\ldots,N_p)$ на итерации t характеризуется четырьмя величинами:

– Положение частицы $\boldsymbol{\alpha}_i^{(t)} = \left[\alpha_{i,1}^{(t)},\ldots,\alpha_{i,D}^{(t)}\right]^T \in \mathbb{R}^D$, где $D = J+1$ — размерность пространства поиска. Вектор $\boldsymbol{\alpha}_i^{(t)}$ представляет текущий набор энергетических порогов.

– Скорость частицы $\mathbf{v}_i^{(t)} = \left[v_{i,1}^{(t)},\ldots,v_{i,D}^{(t)}\right]^T \in \mathbb{R}^D$, определяющая направление и величину изменения положения.

– Индивидуально лучшее положение частицы $\mathbf{p}_i^{(t)} \in \mathbb{R}^D$, соответствующее наилучшему значению целевой функции $F(\boldsymbol{\alpha})$, данной частицей за всё время работы алгоритма вплоть до итерации t.

– Глобально лучшее положение роя: $\mathbf{g}^{(t)} \in \mathbb{R}^D$, при котором значение целевой функции $F(\boldsymbol{\alpha})$ является наилучшим среди всех частиц роя на итерации t.

Структура предложенного алгоритма оптимизации энергетических порогов на основе алгоритма PSO представлена на рисунке 3.4. Алгоритм PSO состоит из следующих шагов:

Шаг 1. Инициализация. На этапе инициализации алгоритма $t = 0$ положения и скорости всех частиц инициализируются случайным образом в пределах заранее

заданных границ пространства поиска:

$$\boldsymbol{\alpha}_i^{(0)} \sim U\left(\left[0,1\right]^{J+1}\right), \mathbf{v}_i^{(0)} \sim U\left(\left[-v_{\max}, v_{\max}\right]^{J+1}\right), \tag{3.26}$$

где $U\left(\left[a,b\right]^D\right)$ обозначает равномерное распределение в D-мерном гиперкубе $\left[a,b\right]^D$. Индивидуально лучшие положения частиц на начальном этапе принимаются равными их начальным положениям:

$$\mathbf{p}_i^{(0)} = \boldsymbol{\alpha}_i^{(0)},\ i = 1,2,\ldots,N_p, \tag{3.27}$$

а глобально лучшее положение роя определяется по следующей формуле:

$$\mathbf{g}^{(0)} = \arg\max_{j=1,\ldots,N_p} F\left(\mathbf{p}_j^{(0)}\right). \tag{3.28}$$

Шаг 2. Итеративное обновление. На каждой последующей итерации $t = 1,2,\ldots,I_{\max}$ выполняется обновление коэффициента инерции, скоростей и положений частиц.

– Обновление коэффициента инерции. Для обеспечения баланса между глобальным исследованием пространства на ранних итерациях и точной локальной эксплуатацией найденного решения на поздних коэффициент инерции $w^{(t)}$ убывает линейно:

$$w^{(t)} = w_{\max} - \frac{\left(w_{\max} - w_{\min}\right)t}{I_{\max}}. \tag{3.29}$$

При высоком значении w частицы сохраняют инерцию предыдущего движения и интенсивно исследуют пространство; при низком значении w — концентрируются вблизи найденных оптимумов.

– Обновление скорости. Скорость i-й частицы в d-й координате обновляется согласно стандартному уравнению PSO [57]:

$$v_{i,d}^{(t+1)} = w^{(t)}.v_{i,d}^{(t)} + c_1.r_{1,d}^{(t)}.\left(p_{i,d}^{(t)} - \alpha_{i,d}^{(t)}\right) + c_2.r_{2,d}^{(t)}.\left(g_d^{(t)} - \alpha_{i,d}^{(t)}\right), \tag{3.30}$$

где c_1 и c_2 — когнитивный и социальный коэффициенты, определяющие вес личного и коллективного опыта; $r_{1,d}^{(t)}$, $r_{2,d}^{(t)}$ — независимые случайные числа в интервале $[0,1]$, генерируемые на каждой итерации для каждой координаты. Первое слагаемое в (3.30) обеспечивает инерцию движения; второе привлекает частицу к

её собственному историческому оптимуму; третье — к глобальному оптимуму роя. Компонента скорости ограничивается: $\left|v_{i,d}^{(t+1)}\right| \le v_{\max}$.

– Обновление положения. Новое положение частицы i в измерении d определяется как:

$$\alpha_{i,d}^{(t+1)} = \alpha_{i,d}^{(t)} + v_{i,d}^{(t+1)}. \tag{3.31}$$

Для обеспечения соблюдения ограничений [0,1] применяется гибридная стратегия обработки границ. Если после обновления $\alpha_{i,d}^{(t+1)} < 0$ или $\alpha_{i,d}^{(t+1)} > 1,$ то:

$$\alpha_{i,d}^{(t+1)} \leftarrow \begin{cases} 0, \text{ если } \alpha_{i,d}^{(t+1)} < 0, \\ 1, \text{ если } \alpha_{i,d}^{(t+1)} > 1, \\ \alpha_{i,d}^{(t+1)}, \text{ иначе,} \end{cases} \tag{3.32}$$

и соответствующая компонента скорости меняет знак с коэффициентом демпфирования β: $v_{i,d}^{(t+1)} \leftarrow -\beta . v_{i,d}^{(t+1)}$, где $\beta = 0{,}5$.

Данная стратегия предотвращает накопление частиц на границах пространства поиска — эффект, характерный для чисто обрезающих стратегий.

– Обновление индивидуально и глобально лучших положений:

Если новое положение частицы приводит к улучшению значения целевой функции, её индивидуально лучшее положение обновляется:

$$\mathbf{p}_i^{(t+1)} = \begin{cases} \boldsymbol{\alpha}_i^{(t+1)}, \text{ если } f\left(\boldsymbol{\alpha}_i^{(t+1)}\right) < f\left(\mathbf{p}_i^{(t)}\right), \\ \mathbf{p}_i^{(t)}, \text{ иначе.} \end{cases} \tag{3.33}$$

Глобально лучшее положение роя обновляется как наилучшее среди всех индивидуально лучших положений:

$$\mathbf{g}^{(t+1)} = \arg \max_{j=1,\ldots,N_p} F\left(\mathbf{p}_j^{(t+1)}\right). \tag{3.34}$$

Шаг 3. Завершение. Алгоритм завершается при выполнении одного из двух условий: достижении максимального числа итераций $t = I_{\max}$, либо при выполнении критерия ранней остановки — отсутствии улучшения глобального оптимума $\mathbf{g}^{(t)}$ в течение заданного числа последовательных итераций при изменении целевой функции менее порогового значения ε.

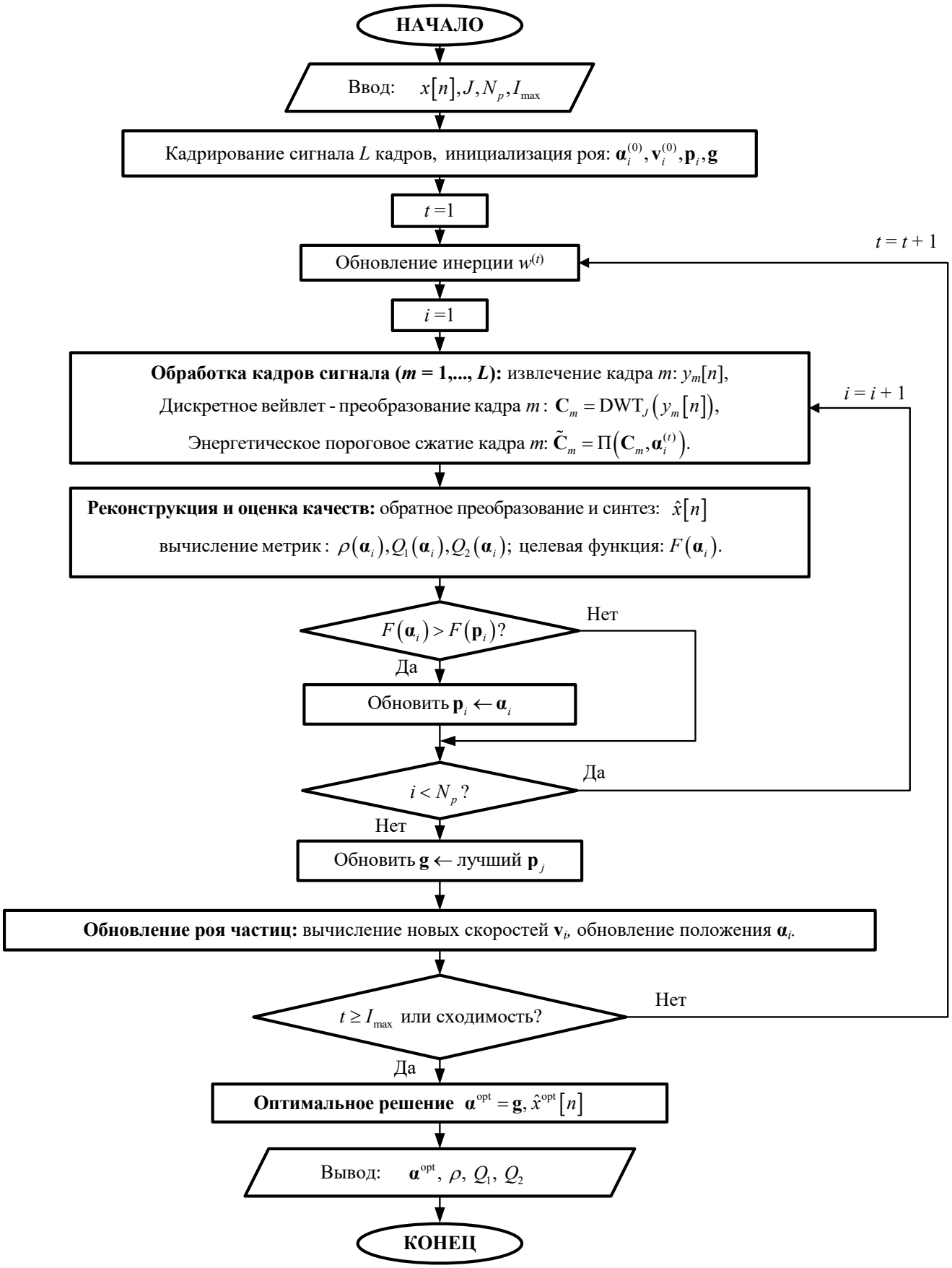


Рисунок 3.4 – Блок-схема алгоритма оптимизации энергетических порогов на основе алгоритма роя частиц PSO

Решением задачи (3.24) считается $\boldsymbol{\alpha}^{\text{opt}} = \mathbf{g}^{(I^*)}$, где I^* — номер завершающей итерации. Этот вектор содержит оптимальные значения энергетических порогов $\alpha_l^{opt} = g_l^{(I^*)}$ для каждого уровня разложения $l = 1,2,\ldots,J$, обеспечивающие максимальную степень сжатия при соблюдении заданных минимальных порогов качества $Q_1^{\min}$ и $Q_2^{\min}$.

3.3.3 Экспериментальная верификация метода: выбор оптимального вейвлет-базиса и результаты оптимизации энергетических порогов

Экспериментальное исследование проведено в два последовательных этапа. Цель первого — выбор вейвлет-базиса и числа уровней разложения, обеспечивающих наилучший потенциал компрессии при фиксированных порогах; полученные результаты служат входными параметрами второго этапа, на котором верифицируется предложенный метод адаптивной оптимизации на основе алгоритма роя частиц PSO. Разделение исследования на два этапа позволяет изолировать вклад вейвлет-базиса от вклада стратегии порогового сжатия и тем самым однозначно атрибутировать прирост эффективности каждому из компонентов системы.

Оба этапа выполнены на одной и той же выборке из 100 речевых записей с частотой дискретизации 8 кГц, отобранных в соответствии с требованиями ГОСТ Р 50840–95 [73]; длина каждой записи не менее 15 секунд, выборка охватывает дикторов обоих полов с различными акустическими характеристиками голоса.

Предварительный этап: выбор типа вейвлета и числа уровней разложения

Алгоритм роя частиц PSO оптимизирует вектор $\boldsymbol{\alpha}$ при фиксированных вейвлет-базисе и числе уровней *J*. Следовательно, выбор пары (вейвлет, *J*) является необходимым предварительным условием: неоптимальный базис задаёт верхнюю границу достижимой разреженности, которую никакая стратегия порогового сжатия не может превысить. Для исследования выбраны 17 вейвлетов четырёх семейств, широко применяемых при обработке РС [83–90]: Добеши (db2, db4, db6, db8, db10); Симлеты (sym4, sym6, sym8, sym10); Койфлеты (coif2, coif3, coif4, coif5); биортогональные вейвлеты (bior2.2, bior3.3, bior4.4, bior5.5).

Теоретически, для сигнала длины M максимальное число уровней разложения определяется как:

$$J_{\max} = \log_2\left(\frac{M}{L_h - 1}\right), \tag{3.35}$$

где L_h — длина фильтра анализа. Для $M = 256$ и типичных длин фильтров $L_h = 4\ldots20$ получем $J_{\max} = 6\ldots7$. Однако практически полезное число уровней ограничено: при малых J энергия недостаточно концентрируется в низкочастотных коэффициентах; при больших J число коэффициентов аппроксимации cA_J становится слишком малым, что приводит к потере информации. Исследован диапазон $J \in \{2,3,4,5,6\}$. Значения $J = 1$ и $J \geq 7$ исключены как заведомо неоптимальные на основе предварительных экспериментов.

Для каждой комбинации (вейвлет, J), где $J \in \{2,3,4,5,6\}$, выполнялось J-уровневое ДВП с последующим энергетическим пороговым сжатием при единых фиксированных порогах $\alpha_l = 0{,}9$ для всех уровней $l = 0,1,\ldots,J$. Применение фиксированного порога исключает влияние стратегии сжатия и позволяет сравнивать базисы исключительно по способности концентрировать энергию речевого сигнала в малом числе коэффициентов. Эффективность каждого вейвлета оценивалась по интегральному критерию (3.24) с весовыми коэффициентами $\mathrm{w}_1 = 0{,}5, \mathrm{w}_2 = 0{,}25, \mathrm{w}_3 = 0{,}25$, реализующими сбалансированный подход между степенью разреженности и качеством восстановления речевого сигнала:

$$F_0 = 0{,}5 f_\rho + 0{,}25 f_{Q_1} + 0{,}25 f_{Q_2}. \tag{3.36}$$

В отличие от целевой функции $F(\boldsymbol{\alpha})$ в формуле (3.25), применяемой при оптимизации методом роя частиц PSO, критерий F_0 в формуле (3.36) не приоритизирует разреженность; равные веса для $Q_1(\boldsymbol{\alpha})$ и $Q_2(\boldsymbol{\alpha})$ обеспечивают, что выбор базиса не смещён в пользу сжатия в ущерб качеству. Для каждого вейвлета определялось оптимальное число уровней:

$$J_{opt} = \arg\max_{J \in \{2,3,\ldots,6\}} F_0(\text{вейвлет}, J), \tag{3.37}$$

после чего вейвлеты сравнивались по максимально достигнутым значениям $F_0\left(\text{вейвлет}, J_{opt}\right)$. На рисунке 3.5 представлены зависимости интегрального критерия эффективности F_0 от числа уровней вейвлет-разложения J для четырёх репрезентативных вейвлетов, принадлежащих различным семействам: Добеши (кривая 1 — db4), Симлеты (кривая 2 — sym6), Койфлеты (кривая 3 — coif3) и биортогональных вейвлетов (кривая 4 — bior3.3).

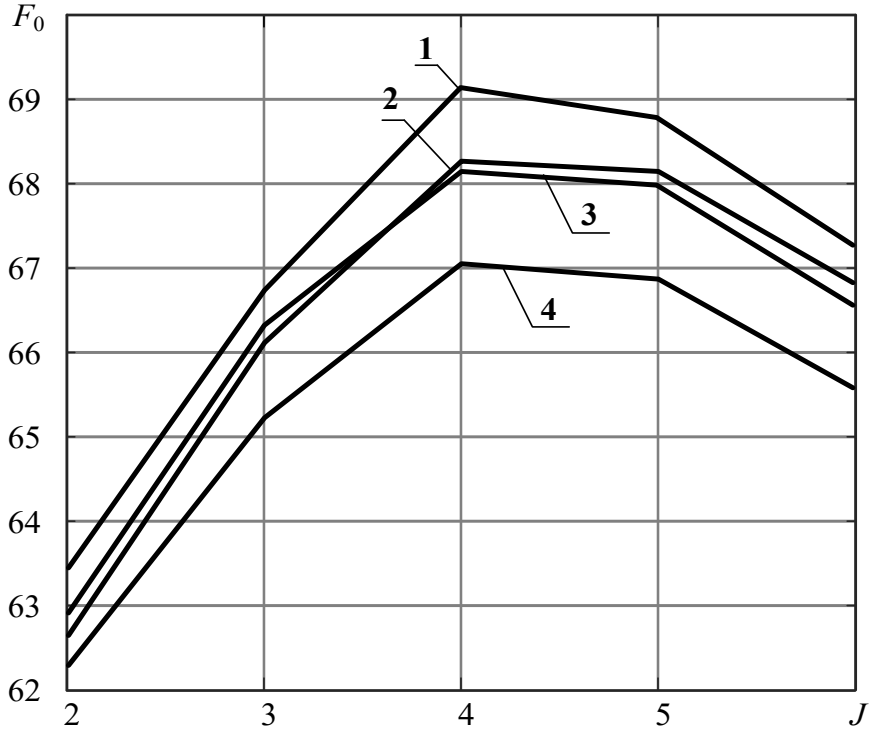


Рисунок 3.5 – Зависимости интегрального критерия эффективности F_0 от числа уровней вейвлет-разложения J для различных типов вейвлетов

Из рисунка 3.5 следует, что для всех исследованных вейвлетов оптимальное число уровней $J_{opt}=4$ — и этот результат инвариантен к выбору семейства. Данный факт не является тривиальным: он указывает на то, что оптимальная глубина разложения определяется прежде всего спектральной структурой самого речевого сигнала, а не свойствами фильтров анализа. Физически объяснение этого эффекта таково: при $M=256$ отсчётах кадра разложение на $J=4$ уровня даёт $L_0=16$ аппроксимирующих коэффициентов, покрывающих полосу 0…500 Гц, где сосредоточена основная просодическая и форматная информация речи; при $.J=3.$ полоса нижнего подпространства вдвое шире (0…1000 Гц), что недостаточно для

разделения второго и третьего формантов; при $J > 4$ аппроксимирующее пространство содержит лишь 8 или 4 коэффициента, что ведёт к деградации восстановления. Снижение F_0 при $J = 2$, $J = 3$ относительно оптимума составляет 3…8 %, при $J > 4$ — 2…4 %. В таблице 3.1 приведено сравнение эффективности сжатия речи при использовании различных вейвлетов при $J = 4$.

Таблица 3.1. Сравнение эффективности сжатия речи при использовании различных вейвлетов при $J = 4$ уровнях разложения

Вейвлет	Коэффициент сжатия (ρ, %)	SegSNR (Q_1, дБ)	ViSQOL (Q_2, балл)	F_0
db2	79,84	11,02	3,72	67,70
db4	81,23	11,34	3,81	69,11
db6	80,95	11,28	3,79	68,83
db8	80,45	11,15	3,75	68,27
db10	79,92	11,05	3,71	67,72
sym4	80,89	11,25	3,78	68,72
sym6	80,52	11,19	3,76	68,38
sym8	80,21	11,12	3,74	68,07
sym10	79,78	11,01	3,70	67,56
coif2	80,68	11,22	3,77	68,54
coif3	80,28	11,14	3,75	68,17
coif4	79,85	11,06	3,72	67,74
coif5	79,42	10,95	3,68	67,23
bior2.2	79,52	10,98	3,69	67,36
bior3.3	79,18	10,89	3,67	67,02
bior4.4	78,95	10,82	3,65	66,74
bior5.5	78,62	10,75	3,63	66,42

Из таблицы 3.1 следует, что при одинаковых уровнях разложения $J = 4$ вейвлет Добеши db4 демонстрирует наилучшие значения интегрального критерия $F_0 = 69{,}11$. Преимущество db4 наблюдается по всем использованным показателям:

коэффициент степень сжатия составляет ρ = 81,23% (максимальное значение), SegSNR достигает $Q_1 = 11{,}34$ дБ, а показатель ViSQOL $Q_2 = 3{,}81$ балла, что также является наилучшим результатом среди сравниваемых вейвлетов.

На основании результатов предварительного этапа установлены следующие параметры ДВП для дальнейшего исследования: вейвлет Добеши db4 и $J = 4$ уровня разложения.

Основной этап: верификация метода оптимизации

На основном этапе выполнена экспериментальная верификация предложенного метода адаптивной оптимизации энергетических порогов на основе алгоритма роя частиц PSO с использованием выбранных параметров вейвлета Добеши db4 и $J = 4$ уровней разложения. Для экспериментальной верификации предложенного метода проведены численные исследования на речевых данных с целью определения оптимальных параметров алгоритма роя частиц PSO, анализа его сходимости и сравнения с существующими методами порогового сжатия на основе ДВП.

Эксперименты выполнены на той же выборке, что и в предварительном исследовании, — 100 записей РС, отобранных в соответствии с требованиями стандарта ГОСТ Р 50840–95.

Параметры алгоритма PSO установлены следующим образом:

– Размер роя установлен $N_p = 200$ частиц, что обеспечивает достаточное покрытие пятимерного пространства поиска $\left(D = J + 1 = 5\right)$ при приемлемой вычислительной сложности.

– Максимальное число итераций $I_{\max} = 100$.

– Коэффициент инерции изменяется линейно от $w_{\max} = 0{,}9$ до $w_{\min} = 0{,}3$, обеспечивая баланс между глобальным поиском на начальном этапе и локальной оптимизацией на завершающем.

– Когнитивный и социальный коэффициенты установлены равными $c_1 = c_2 = 2{,}0$ согласно стандартной параметризации PSO.

– Максимальная скорость ограничена значением $v_{\max} = 0{,}15$ (15% от диапазона

поиска) для предотвращения чрезмерно больших скачков в пространстве решений.

– Коэффициент демпфирования при отражении $\beta = 0{,}5$.

– Область поиска для всех компонент вектора порогов установлена как $\alpha_l \in [0,1]$, где $l = 0,1,\ldots,J = 4$.

– Для целевой функции использован режим с приоритетом высокой степени сжатия, заданный весовыми коэффициентами $w_1 = 0{,}7, w_2 = 0{,}15, w_3 = 0{,}15$ при минимально допустимом качестве восстановленного РС. В качестве ограничений на качество восстановленного сигнала приняты минимально допустимые значения: $Q_1(\boldsymbol{\alpha}) \geq Q_1^{\min} = 9{,}5$ дБ (по SegSNR) и $Q_2(\boldsymbol{\alpha}) \geq Q_2^{\min} = 3{,}5$ баллов (по шкале MOS), в соответствии с рекомендациями ITU-T P.800 [72].

– Критерий ранней остановки: алгоритм завершается при отсутствии улучшения глобального наилучшего решения $F(\mathbf{g}^{(t)})$ в течение 20 последовательных итераций при изменении целевой функции менее $\varepsilon = 10^{-6}$. Данный критерий позволяет сократить вычислительное время для записей с быстрой сходимостью без потери качества решения. Нижняя граница $\alpha_l \geq 0{,}1$ гарантирует сохранение не менее 10 % энергии на каждом уровне разложения, что предотвращает полное подавление уровней с непренебрежимым речевым содержанием.

На рисунке 3.6 представлены результаты многокритериальной оптимизации энергетических порогов вейвлет-преобразования с использованием алгоритма роя частиц PSO, отражающие изменение ключевых показателей эффективности по итерациям.

На рисунке 3.6а показана зависимость коэффициента сжатия ρ от номера итерации *t*. На начальном этапе (итерации 0...10) наблюдается интенсивный рост ρ от 81% до 86%, затем темп замедляется (итерации 11...35), и с итерации 36 устанавливается стационарный режим $\rho = 87\%$; алгоритм останавливается на 48-й итерации по критерию ранней остановки, расходуя лишь 48 % от максимального числа итераций $I_{\max} = 100$; это подтверждает эффективность критерия сходимости и указывает на отсутствие необходимости дальнейшего продолжения поиска.

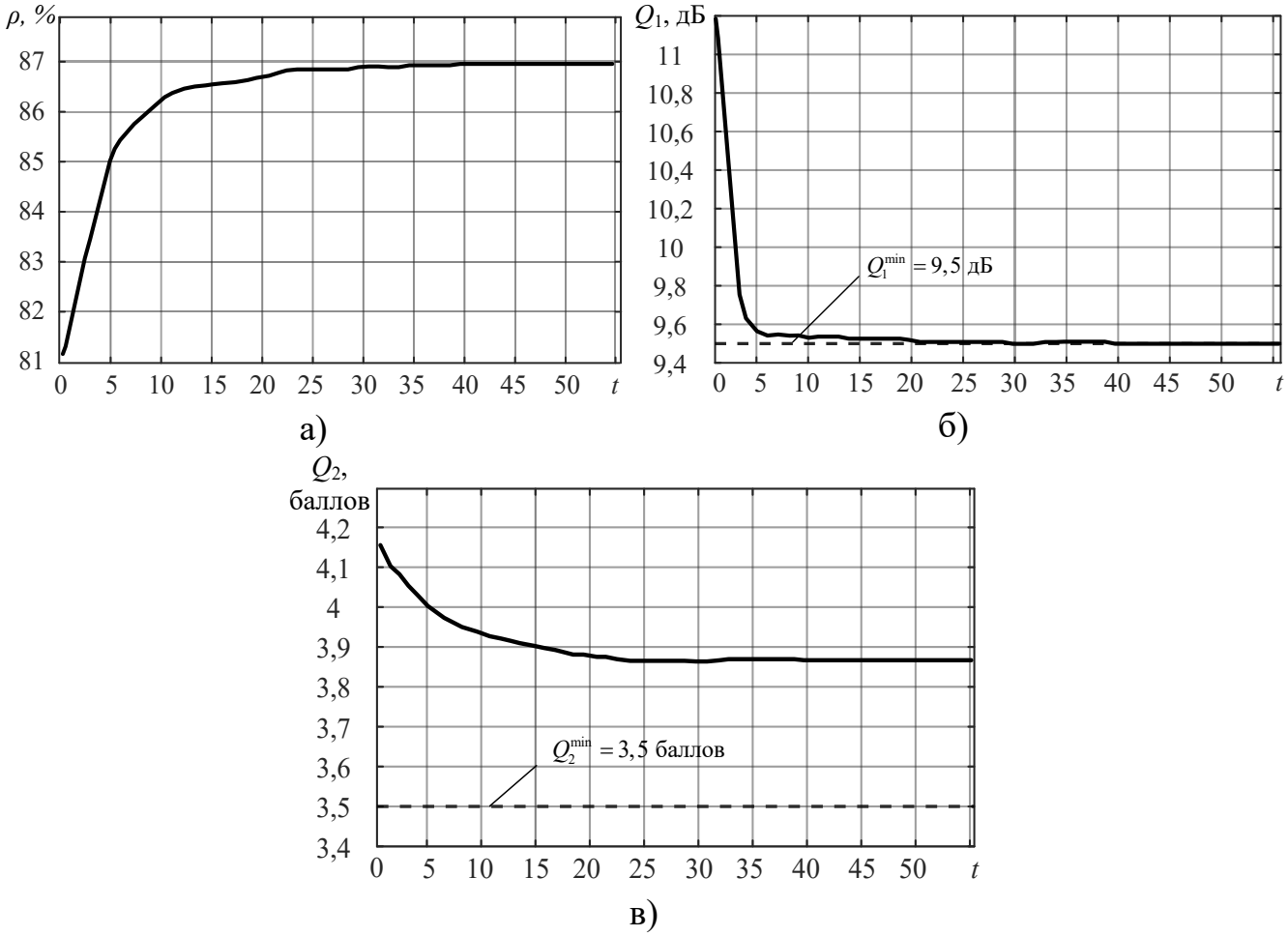


Рисунок 3.6. Результаты многокритериальной оптимизации методом PSO: зависимости от номера итерации t

(а) коэффициента сжатия; (б) SegSNR; (в) ViSQOL

На рисунке 3.6б представлена зависимость SegSNR от t с горизонтальной линией минимального требования $Q_1^{\min} = 9{,}5$ дБ. После начального превышения ограничения (11,2 дБ) метрика снижается и стабилизируется на уровне 9,50±0,05 дБ (итерации 13...55), точно соответствуя минимальному требованию.

На рисунке 3.6в показана эволюция перцептуального качества ViSQOL. Начальное значение ViSQOL = 4,18 (субъективно «хорошо» по шкале MOS) снижается к итерации 5 до 4,0 под действием целевой функции, которая приоритизирует рост ρ при допустимом падении качества. После итерации 25 ViSQOL стабилизируется на уровне 3,87±0,01, что соответствует оценке «удовлетворительно–хорошо» и обеспечивает запас качества речи $\Delta Q_2 = 0{,}37$ относительно минимального требования $Q_2^{\min} = 3{,}5$ балла.

Принципиально важное наблюдение: ограничение по Q_1 оказывается активным в точке оптимума — алгоритм находит решение точно на границе допустимого множества $\left(Q_1 = Q_1^{\min}\right)$, тогда как ограничение по Q_2 остаётся неактивным с запасом $\Delta Q_2 = 0{,}37$ балла. Это означает, что Q_1 является лимитирующим критерием качества для задачи максимизации разреженности речевого сигнала в данных условиях.

По завершении оптимизации алгоритм роя частиц PSO сходится к вектору: $\boldsymbol{\alpha}^{\text{opt}} = \left[0{,}989\ 0{,}954\ 0{,}584\ 0{,}447\ 0{,}171\right]^T$. Это означает, что для обеспечения высокого качества восстановления речи сохраняется почти вся энергия на низкочастотных уровнях (аппроксимация и первые детали), в то время как на высокочастотных уровнях допускается сильное сжатие. Такое распределение полностью отражает специфику РС — концентрацию основной энергии в низкочастотной области — и обеспечивает оптимальный баланс между степенью сжатия и разборчивостью. Для оценки эффективности проведено сравнение с классическими методами. В таблице 3.2 приведено сравнение предложенного метода PSO с тремя базовыми подходами к пороговой обработке вейвлет-коэффициентов, усреднённое по выборке из 100 РС. Для всех методов использовались идентичные параметры вейвлет-преобразования.

Анализ результатов, представленных в таблице 3.2, позволяет сделать следующие выводы. Метод универсального порога Донохо – Джонстона нарушает ограничение $Q_1^{\min}$: при $Q_1 = 9{,}18$ дБ нарушение составляет 0,32 дБ, что делает метод неприменимым при заданных требованиях к качеству; причина — метод разработан для гауссовского шума и не учитывает тяжёлые хвосты распределения вейвлет-коэффициентов на энергетически насыщенных уровнях речевого сигнала.

Метод SUREShrink соблюдает ограничения, однако оставляет значительный запас по $Q_1 = 10{,}81$ дБ (превышение $Q_1^{\min}$ на 1,31 дБ), что свидетельствует о консервативной стратегии, не использующей полностью допустимое пространство сжатия.

Таблица 3.2. Сравнительный анализ методов пороговой обработки вейвлет-коэффициентов для сжатия РС

Метод	Коэффициент сжатия (ρ, %)	SegSNR (Q_1, дБ)	ViSQOL (Q_2, балл)
Универсальный порог Донохо – Джонстона	76,32	9,18	3,45
Адаптивный SUREShrink	79,47	10,81	3,68
Энергетический порог с фиксированными значениями $\alpha_l = 0{,}9$	81,23	11,34	3,81
Энергетический порог, оптимизированный с помощью алгоритма PSO	87,00	9,5	3,87

Энергетический метод с фиксированным порогом $\alpha_l = 0{,}9$ обеспечивает высокое качество ($Q_1 = 11{,}34$ дБ), но с запасом 1,84 дБ — именно этот запас алгоритм PSO конвертирует в прирост разреженности.

Предложенный PSO-метод достигает $\rho = 87{,}0\%$ при точном выходе $Q_1 = 9{,}5$ дБ на границу ограничения, с приростом разреженности +14,7% к методу Донохо–Джонстона, +7,5% к SUREShrink и +5,8% к фиксированному энергетическому порогу. Примечательно, что $Q_2 = 3{,}87$ балла у предложенного метода превышает значения всех конкурентов: дифференцированное подавление высокочастотных уровней, реализованное через $\boldsymbol{\alpha}^{\text{opt}}$, сохраняет перцептуально значимые составляющие речи точнее, чем равномерная фиксированная стратегия. Полученный вектор $\boldsymbol{\alpha}^{\text{opt}}$ и подтверждённая эффективность метода создают основу для его применения в рамках интегрированной системы совместной обработки, исследуемой в разделе 3.4. При переходе к двухканальной архитектуре на базе представления Хургина – Яковлева размерность пространства поиска удваивается до $D = 2(J+1) = 10$, что потребует совместной оптимизации порогов обоих каналов с сохранением всех параметров алгоритма PSO, верифицированных в настоящем разделе.

3.4 Исследование интегрированной системы кодирования речи на основе представления Хургина – Яковлева и дискретного вейвлет-преобразования

3.4.1 Структура и параметры интегрированной системы

Результаты, полученные в разделах 3.2 и 3.3, показывают, что ДВП с покадровым энергетическим пороговым сжатием, оптимизируемым алгоритмом PSO, обеспечивает степень разреженности ρ до 87,0 % при соблюдении ограничений на инструментальное и перцептивное качество восстановленной речи. В то же время из результатов главы 2 следует, что система на основе представления Хургина – Яковлева при числе каналов $K = 2$ и числе этапов разложения $P = 1$ обеспечивает выигрыш качества ViSQOL до 0,77 балла по шкале MOS по сравнению с эталонной системой на основе теоремы В.А. Котельникова — прежде всего за счёт повышения помехоустойчивости и декорреляции субполосных сигналов.

Указанные два метода решают взаимодополняющие задачи. Представление Хургина – Яковлева формирует компактное многоканальное разложение, снижающее межотсчётную избыточность исходного сигнала, тогда как ДВП с пороговым сжатием устраняет внутрикадровую избыточность в пространстве субполосных сигналов. Объединение этих механизмов в единой архитектуре открывает возможность получить синергетический эффект, при котором декорреляция, выполненная блоком представления Хургина – Яковлева, улучшает вейвлет-разреженность субполосных сигналов и тем самым повышает эффективность последующего порогового сжатия. Настоящий раздел посвящён исследованию такой интегрированной системы.

Структурная схема интегрированной системы первичного кодирования РС представлена на рисунке 3.7.

На передающей стороне РС после АЦП поступает в блок анализа представления Хургина – Яковлева, реализующий K-канальное P-этапное разложение. В результате формируются K^P субполосных сигналов $y_k[n], k = 0,1,\ldots,K^P - 1,$ каждый из которых имеет частоту дискретизации $f_Д / K^P$ и несёт информацию о соответствующей производной исходного сигнала. Каждый

субполосный сигнал $y_k[n]$ разбивается на L непересекающиеся кадры $y_k^{(m)}[n]$ длиной $M = 256$ отсчётов, $m = 1,\ldots,L$. К каждому кадру независимо применяется J-уровневое ДВП по быстрому алгоритму Малла:

$$\mathbf{C}_k^{(m)} = \mathrm{DWT}_J\left(y_k^{(m)}[n]\right),\ k = 0,1,\ldots,K^P - 1,\ m = 1,\ldots,L, \tag{3.38}$$

где $\mathbf{C}_k^{(m)}$ — вектор всех вейвлет-коэффициентов k-го канала m-го кадра, содержащий коэффициенты аппроксимации $cA_J^{(m,k)}$ и коэффициенты деталей $cA_j^{(m,k)}, j = 1,\ldots,J$:

$$\mathbf{C}_k^{(m)} = \left[cA_J^{(m,k)}; cD_J^{(m,k)}; cD_{J-1}^{(m,k)}; \ldots; cD_1^{(m,k)}\right]^T, \tag{3.39}$$

общей длиной M (при периодических граничных условиях). Скалярный коэффициент i-й позиции уровня l, кадра m, канала k обозначается $c_i^{(l,m,k)}$.

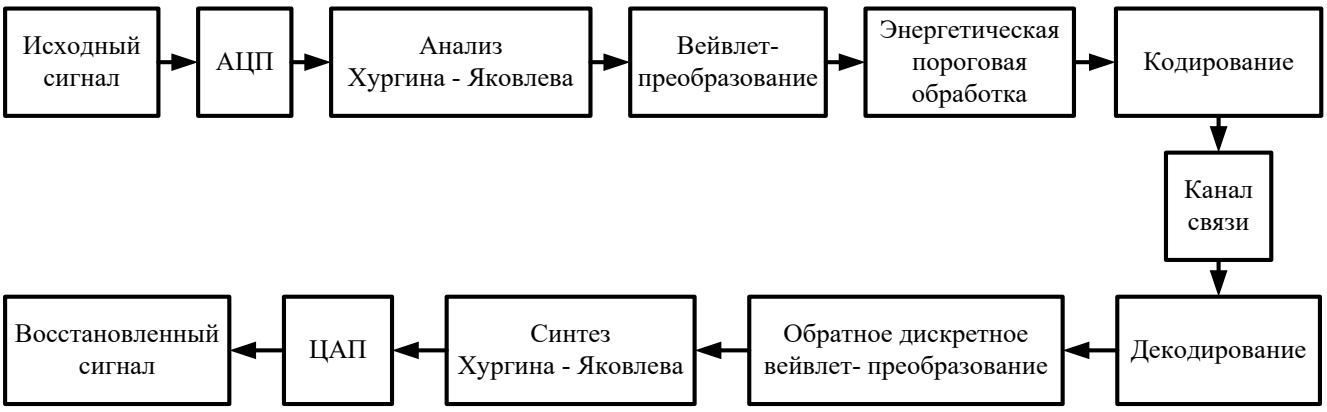


Рисунок 3.7 – Структурная схема интегрированной системы обработки и передачи сигналов на основе представления Хургина – Яковлева и вейвлет-преобразования

К вейвлет-коэффициентам каждого уровня l каждого канала k применяется энергетическое пороговое сжатие с параметром $\alpha_{k,l}$, формируя вектор $\tilde{\mathbf{C}}_k^{(m)}$ с обнулёнными незначимыми коэффициентами. Сжатые векторы передаются по каналу связи. Полный вектор энергетических порогов интегрированной системы задаётся с двойной индексацией по каналу k и уровню l:

$$\boldsymbol{\alpha} = \left[\alpha_{0,0}, \alpha_{0,1}, \ldots, \alpha_{0,J}, |\, \alpha_{1,0}, \alpha_{1,1}, \ldots, \alpha_{1,J}, |\, \ldots, |\, \alpha_{K^P-1,0}, \ldots, \alpha_{K^P-1,J}\right]^T, \tag{3.40}$$

где размерность вектора порогов определяется по формуле:

$$D = K^P(J+1). \tag{3.41}$$

На стороне приёмника последовательно выполняются: ОДВП для каждого

канала k и каждого кадра m:

$$\hat{y}_k^{(m)}[n] = \mathrm{IDWT}_J\left(\tilde{\mathbf{C}}_k^{(m)}\right), \tag{3.42}$$

а затем синтез сигнала блоком представления Хургина – Яковлева с интерполяцией и синтезирующими фильтрами, представленными в главе 1. После восстановления цифровых отсчётов сигнал поступает на цифро-аналоговый преобразователь (ЦАП) для формирования аналогового речевого сигнала. Принципиальная особенность данной архитектуры состоит в том, что блоки ДВП/ОДВП встроены между анализирующим и синтезирующим каскадами представления Хургина – Яковлева и функционируют в пространстве субполосных сигналов, а не в пространстве исходного сигнала. Это позволяет адаптировать стратегию порогового сжатия к статистическим свойствам каждого субполосного канала независимо.

Выбор конкретных значений параметров интегрированной системы обработки определяется совокупностью требований: обеспечение точного восстановления в дискретной области, минимизация накапливаемых ошибок, реализуемость в условиях ограниченной вычислительной мощности, а также максимизация выигрыша в качестве восстановленной речи по критериям SegSNR и ViSQOL.

– Выбор числа каналов $K = 2$ и числа этапов $P = 1$. Из анализа главы 1 (таблица 1.1) следует, что конфигурация $K = 2$, $P = 1$ обеспечивает исключительно высокое инструментальное качество восстановления речи: SNR = 85,38 дБ при MSE = $7{,}6 \cdot 10^{-12}$ и ViSQOL = 4,73 балла. Принципиальное математическое основание данного выбора состоит в следующем. Для двухканальной системы ($K = 2$) система уравнений синтезирующих фильтров в непрерывной области имеет аналитическое решение в замкнутой форме с ненулевым определителем, обеспечивая точное восстановление непрерывного сигнала. Для K = 3 и K = 4 замкнутое решение в непрерывной области недостижимо: применяется кусочно-линейная аппроксимация по критерию минимума СКО, вносящая погрешность $\varepsilon_a > 0$ в АЧХ синтезирующих фильтров.

При $K = 2$, $P = 1$ суммарная ошибка восстановления обусловлена исключительно точностью вычисления первой производной $\left(\varepsilon_{der} \le 0{,}01\right)$, тогда как

погрешность аппроксимации АЧХ синтезирующих фильтров полностью отсутствует $(\varepsilon_a = 0)$. Многоэтапное разложение $(P > 1)$ при $K = 2$ приводит к накоплению производных погрешностей: SNR снижается с 85,38 дБ до 55,93 дБ при $P = 2$ и до 48,88 дБ при $P = 3$. Таким образом, конфигурация $K = 2$, $P = 1$ является оптимальной с точки зрения баланса точности восстановления, выигрыша помехоустойчивости и вычислительной сложности.

– Выбор типа вейвлета и числа уровней разложения $J = 4$. В разделе 3.3 проведено систематическое сравнение 17 вейвлет-базисов четырёх семейств по критерию максимизации целевой функции F_0 при фиксированных ограничениях качества. Вейвлет db4 (Добеши 4-го порядка, $J = 4$) обеспечил наилучший результат. Оптимальность $J = 4$ была установлена экспериментально и оказалась инвариантной к семейству вейвлетов. Указанные параметры применяются к каждому из $K = 2$ каналов интегрированной системы.

Таким образом, итоговая конфигурация интегрированной системы определяется параметрами: число каналов анализа на основе представления Хургина – Яковлева $K = 2$, число этапов $P = 1$, тип вейвлета db4, число уровней ДВП $J = 4$. Полная размерность вектора порогов составляет: $D = K^P(J+1) = 2(4+1) = 10$ и вектор индексируется как $\boldsymbol{\alpha} = \left[\alpha_{0,0}, \alpha_{0,1}, \alpha_{0,2}, \alpha_{0,3}, \alpha_{0,4}, \alpha_{1,0}, \alpha_{1,1}, \alpha_{1,2}, \alpha_{1,3}, \alpha_{1,4}\right]^T$, где первый индекс $(k = 0,1)$ обозначает номер канала (сигнал и первая производная соответственно), а второй $(l = 0,\ldots,4)$ — номер уровня ДВП.

3.4.2 Оптимизация параметров интегрированной системы методом роя частиц PSO

Задача оптимизации вектора порогов $\boldsymbol{\alpha} \in [0,1]^{10}$ интегрированной системы формулируется как многокритериальная задача максимизации коэффициента сжатия при ограничениях на качество восстановления речи:

$$\boldsymbol{\alpha}^{\text{opt}} = \arg\max_{\boldsymbol{\alpha}\in[0,1]^{10}} F(\boldsymbol{\alpha}) = \arg\max_{\boldsymbol{\alpha}\in[0,1]^{10}} \left[\mathrm{w}_1 f_\rho(\boldsymbol{\alpha}) + \mathrm{w}_2 f_{Q_1}(\boldsymbol{\alpha}) + \mathrm{w}_3 f_{Q_2}(\boldsymbol{\alpha}) \right], \qquad (3.43)$$

при ограничениях $Q_1(\boldsymbol{\alpha}) \geq Q_1^{\min} = 9{,}5$ дБ (по SegSNR) и $Q_2(\boldsymbol{\alpha}) \geq Q_2^{\min} = 3{,}5$ балла (по шкале MOS), где целевая функция $F(\boldsymbol{\alpha})$, нормализованные метрики качества $f_{Q_1}(\boldsymbol{\alpha}), f_{Q_2}(\boldsymbol{\alpha})$ и весовые коэффициенты $\mathrm{w}_1 = 0{,}7$, $\mathrm{w}_2 = \mathrm{w}_3 = 0{,}15$ определены в разделе 3.3. В отличие от одноканальной задачи, задача имеет увеличенную размерность поискового пространства $D = 10$. Алгоритм PSO применяется без модификации параметров: $N_p = 200$ частиц, $I_{\max} = 100$ итераций, коэффициент инерции изменяется линейно от $w_{\max} = 0{,}9$ до $w_{\min} = 0{,}3$, когнитивный и социальный коэффициенты установлены равными $c_1 = c_2 = 2{,}0$, ограничение скорости $v_{\max} = 0{,}15$. Критерий ранней остановки: алгоритм завершается при отсутствии улучшения глобального наилучшего решения $F\left(\mathbf{g}^{(t)}\right)$ в течение 20 последовательных итераций при изменении целевой функции менее $\varepsilon = 10^{-6}$.

Результаты оптимизации представлены на рисунке 3.8.

На рисунке 3.8а показана динамика коэффициента сжатия $\rho(t)$: в течение первых 10 итераций наблюдается интенсивный рост с ~86 % до ~93 %, затем темп замедляется и, начиная с итерации $t \approx 40$ устанавливается стационарный режим $\rho = 92{,}9\%$. Алгоритм завершает работу на итерации $t = 45$. Достигнутое значение $\rho = 92{,}9\%$ превышает результат одноканальной оптимизации на 5,9 процентных пункта, что подтверждает синергетический эффект совместного применения представления Хургина – Яковлева и вейвлет-порогового сжатия.

Динамика $Q_1(t)$ (рисунок 3.8б) демонстрирует процедуру активного ограничения: Q_1 стабилизируется на уровне 9,50 ± 0,05 дБ, точно соответствуя нижнему пределу $Q_1^{\min}$. Ограничение по ViSQOL является неактивным: $Q_2 = 3{,}6$ балла при запасе 0,10 балла относительно $Q_2^{\min} = 3{,}5$ (рисунок 3.8в). Данная структура активных ограничений согласуется с результатами одноканальной оптимизации раздела 3.3.3 и свидетельствует об устойчивости выбранной конфигурации весовых коэффициентов к изменению размерности задачи.

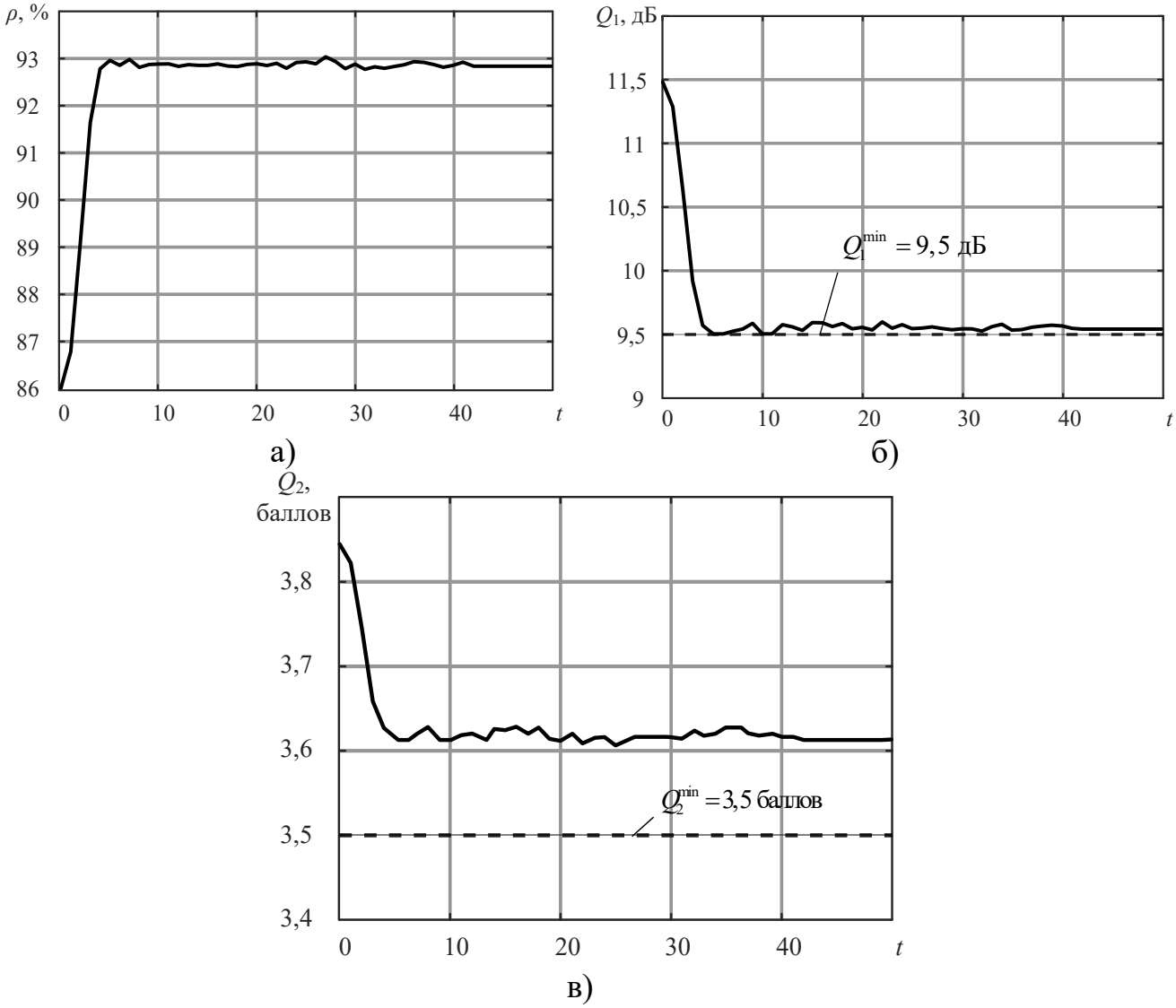


Рисунок 3.8 – Результаты многокритериальной оптимизации методом PSO: зависимости от номера итерации t

(а) коэффициента сжатия; (б) SegSNR; (в) ViSQOL

Оптимальный вектор порогов можно записать в следующем виде:

$$\boldsymbol{\alpha}^{\text{opt}} = \Big[\underbrace{0{,}991 \quad 0{,}856 \quad 0{,}602 \quad 0{,}385 \quad 0{,}178}_{\text{Канал 0 (сигнала)}} \ \underbrace{0{,}987 \quad 0{,}823 \quad 0{,}571 \quad 0{,}362 \quad 0{,}185}_{\text{Канал 1 (производная)}} \Big]^{T}.$$

Анализ оптимального вектора выявляет закономерность, единую для обоих каналов: $\alpha_{k,l}$ убывает монотонно от уровня аппроксимации $(l=0)$ к уровню наибольшей детализации $l=J=4$, отражая фундаментальное свойство РС — концентрацию энергии в низкочастотных подполосах.

Сравнение каналов между собой обнаруживает систематическое расхождение: на уровнях детализации $l=1,\ 2,\ 3$ порог канала 1 (производная) ниже

соответствующего порога канала 0 (сигнал) на 0,03…0,04, тогда как на уровне $l = 4$ наблюдается незначительное превышение (0,185 против 0,178).

Физическая интерпретация этого расхождения следующая. Первая производная речевого сигнала акцентирует переходные процессы — границы фонем, взрывные согласные — и характеризуется более равномерным распределением энергии по уровням разложения по сравнению с исходным сигналом. На промежуточных уровнях детализации $(l = 1\ldots3)$ относительный вклад незначимых коэффициентов в энергию канала 1 выше, чем в канале 0, что позволяет более агрессивно обнулять их без нарушения ограничения $Q_1 \geq Q_1^{\min}$.

На самом тонком уровне ($l = 4$), напротив, производная содержит перцептивно значимые высокочастотные детали — быстрые переходы речевого тракта — и требует несколько более высокой доли сохранённой энергии. Раздельное назначение одинакового вектора порогов обоим каналам не позволяет учесть эту асимметрию: занижение $\alpha_{1,4}$ привело бы к нарушению ограничения Q_1, завышение $\alpha_{1,1\ldots3}$ — к потере 1,5…2,0 процентных пункта. разреженности. Совместная PSO-оптимизация автоматически находит баланс, недостижимый при покаловой настройке.

Прирост коэффициента сжатия интегрированной системы на 5,9 процентных пункта достигается при неизменных ограничениях на качество и обусловлен исключительно структурным эффектом: блок представления Хургина – Яковлева формирует субполосные сигналы с более выраженной вейвлет-разреженностью, чем исходный сигнал, создавая тем самым более благоприятные условия для энергетического порогового сжатия в каждом канале.

3.4.3 Сравнение эффективности интегрированной системы с низкоскоростными кодеками

Для оценки практической эффективности интегрированной системы проведено экспериментальное сравнение трёх систем первичного кодирования речи:

– система 1: кодирование на основе теоремы В.А. Котельникова $(K = 1)$ совместно с кодеком первичного кодирования;

– система 2: кодирование на основе ДВП с PSO-оптимизированными энергетическими порогами совместно с кодеком первичного кодирования (вейвлет db4, $J = 4$, $\boldsymbol{\alpha}^{\text{opt}}$ из раздела 3.3.3);

– система 3 (предложенная): интегрированная система на основе представления Хургина – Яковлева $(K = 2, P = 1)$ и ДВП с PSO-оптимизированным порогами (вейвлет db4, $J = 4$, $\boldsymbol{\alpha}^{\text{opt}}$ из раздела 3.4.3) совместно с кодеком первичного кодирования.

Исследование проводится с использованием кодека LBRAMR (Low-Bitrate AMR, тип ACELP) в диапазоне скоростей передачи $C = 1{,}0 \ldots 5{,}0$ кбит/с, что соответствует условиям работы тактических систем радиосвязи. Вероятность битовых ошибок в канале: $P_{ош} = 0\%,\ 1\%,\ 5\%$ при случайной модели ошибок. Методика эксперимента соответствует разделу 1.6.2. Критерий качества — ViSQOL по шкале MOS.

Результаты представлены на рисунке 3.9. Кривая 1 — система на основе теоремы В.А. Котельникова; кривая 2 — ДВП с PSO-оптимизированными порогами; кривая 3 — интегрированная система на основе представления Хургина – Яковлева и ДВП с PSO-оптимизированными порогами. Анализ результатов позволяет сформулировать следующие выводы.

Все три кривые монотонно возрастают с увеличением C при фиксированной $P_{ош}$, что соответствует общей закономерности: рост скорости передачи снижает степень параметрического квантования в кодеке LBRAMR и тем самым уменьшает искажения восстановленного речевого сигнала. Кривая 3 стабильно расположена выше кривых 1 и 2 во всём исследованном диапазоне условий, что свидетельствует о устойчивом преимуществе интегрированной системы.

При $P_{ош} = 0\%$ (рисунок 3.9а) разрыв между кривыми 1 и 3 максимален при $C = 1{,}0$ кбит/с и постепенно сокращается с ростом C, хотя кривая 3 сохраняет преимущество вплоть до $C = 5{,}0$ кбит/с. При малых скоростях передачи кодек LBRAMR вынужден функционировать в режиме предельной компрессии, при котором доля квантового шума в восстановленном сигнале максимальна, а перцептивный вес точного воспроизведения переходных процессов — границ

фонем и взрывных согласных — наиболее высок. Именно в этих условиях вклад блока представления Хургина – Яковлева, точнее передающего первую производную речевого сигнала и тем самым улучшающего аппроксимацию переходных участков, оказывается максимальным. С ростом C роль данного фактора снижается, поскольку кодек располагает достаточной скоростью для прямой параметрической передачи этих участков без посредства производной.

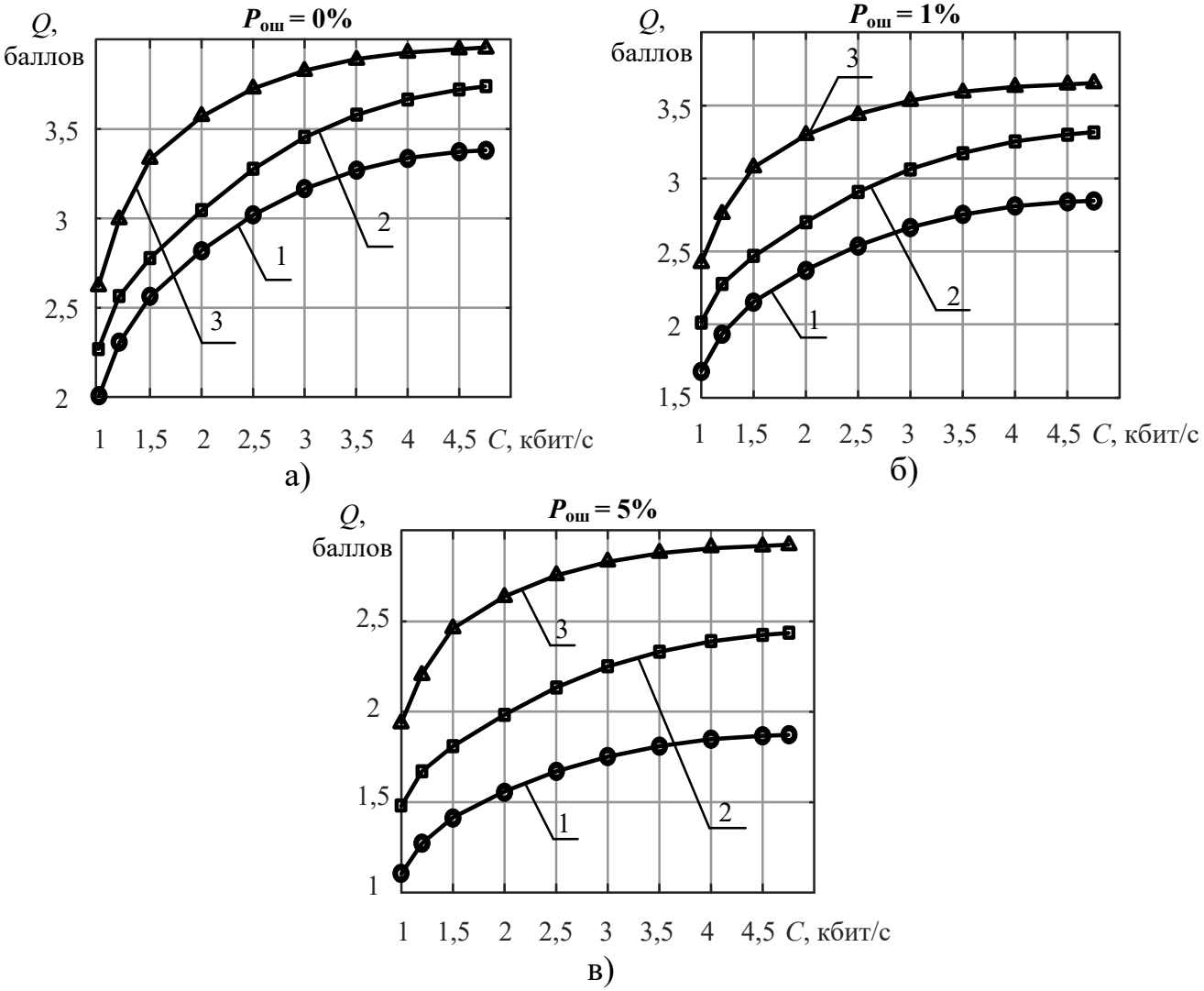


Рисунок 3.9 – Оценка качества восстановленной речи на выходе декодера LBRAMR при различных скоростях передачи C и вероятностях ошибок в канале: (а) $P_{ош} = 0\%$, (б) $P_{ош} = 1\%$, (в) $P_{ош} = 5\%$

Сравнение кривых 2 и 3 при $P_{ош} = 0\%$ выявляет устойчивый положительный разрыв, не исчезающий при увеличении C. Это демонстрирует, что прирост разреженности с 87,0 % до 93,0 %, обеспечиваемый декорреляцией блока представления Хургина – Яковлева, транслируется в измеримое улучшение

качества восстановленного речевого сигнала на выходе декодера кодека — то есть разреженность вейвлет-представления обладает прямой операциональной ценностью для данного класса первичных кодеков.

При переходе от $P_{ош} = 0\%$ к $P_{ош} = 1\%$ (рисунок 3.9б) и $P_{ош} = 5\%$ (рисунок 3.9в) качество всех трёх систем закономерно снижается, однако темп снижения для систем 1 и 2 существенно выше, чем для системы 3. Вследствие этого выигрыш $\Delta Q_1 = Q_3 - Q_1$ возрастает с увеличением $P_{ош}$, что означает: интегрированная система не только превосходит эталонную в условиях идеального канала, но и проявляет более высокую помехоустойчивость.

Данный эффект обусловлен двумя механизмами. Во-первых, блок представления Хургина – Яковлева формирует ортогональную субполосную структуру, в которой ошибочные биты, приходящиеся на один канал, не распространяются на другой, что снижает кумулятивный перцептивный ущерб. Во-вторых, PSO-оптимизированные пороги в системе 3 сосредотачивают максимальное сжатие в высокочастотных вейвлет-подполосах, где перцептивный вес коэффициентов минимален: ошибочные биты, попадающие на обнулённые коэффициенты, не формируют слышимых артефактов. В системе 2 аналогичный механизм ослаблен вследствие меньшей разреженности, а в системе 1 вовсе отсутствует.

Взаимодействие двух указанных механизмов носит синергетический характер: при $P_{ош} = 5\%$ разрыв $Q_3 - Q_2$ превышает разрыв $Q_2 - Q_1$, что свидетельствует о том, что вклад блока представления Хургина – Яковлева в условиях ошибочного канала превышает вклад самого вейвлет-сжатия.

Интегрированная система на основе представления Хургина – Яковлева, ДВП и оптимизации энергетических порогов методом роя частиц обеспечивает выигрыш в качестве восстановленной речи по метрике ViSQOL до 0,9 балла по шкале MOS по сравнению с системой на основе теоремы В.А. Котельникова. Весь наблюдаемый выигрыш достигается без увеличения скорости передачи данных по сравнению с эталонной системой, что подтверждает эффективность предложенного решения именно для низкоскоростных тактических систем радиосвязи.

3.5 Предложения по реализации предложенных систем на современной элементной базе

3.5.1 Анализ вычислительной сложности интегрированной системы

Практическая значимость разработанных алгоритмов определяется не только их теоретической эффективностью, но и возможностью реализации в условиях ограниченных вычислительных ресурсов встраиваемых радиотехнических систем реального времени. Настоящий раздел решает три взаимосвязанные задачи: анализ вычислительной сложности интегрированной системы, оценку требуемой производительности платформы и обоснование рекомендаций по выбору элементной базы.

Все оценки относятся к конфигурации, зафиксированной в разделе 3.4.2: число каналов $K = 2$, число этапов разложения $P = 1$, вейвлет db4, число уровней ДВП $J = 4$, длина кадра $M = 256$ отсчётов, частота дискретизации 8 кГц.

Анализ проводится на двух уровнях. Первый уровень — подсчёт точного числа арифметических операций на кадр при зафиксированных параметрах системы, являющийся непосредственной входной величиной для расчёта загрузки платформы. Второй уровень — асимптотические оценки O(·), устанавливающие порядок роста числа операций при увеличении длины кадра M и определяющие доминирующие блоки при масштабировании системы.

Блок анализ Хургина – Яковлева ($K = 2,\ P = 1$):

Алгоритм включает три вычислительно значимых шага: однократное прямое БПФ(M) для формирования общего спектра всех K каналов; умножение на операторы дифференцирования ($K - 1$) каналов в частотной области; и K обратных ОБПФ(M) до децимации (подробный пошаговый вывод: Приложение А, таблица А.1). Число вычислительных операций блока анализа определяется выражениями:

$$C_{A,\times} = 2(K+1)M\log_2 M + 4(K-1)M \ \text{(умножения)}, \tag{3.44}$$

$$C_{A,+} = 3(K+1)M\log_2 M + 2(K-1)M \ \text{(сложения)}. \tag{3.45}$$

При $M = 256$, $K = 2$: $C_{A,\times} = 13312$ (умножения), $C_{A,+} = 18944$ (сложения), получим полное число вещественных арифметических операций блока анализа

Хургина – Яковлева за один кадр:

$$C_A = C_{A,+} + C_{A,\times} = 18944 + 13312 = 32256 \text{ оп./кадр.}$$

Из структуры формулы (3.44), (3.45) следует, что при фиксированном K число операций блока анализа растёт как $\mathrm{O}\left(M \log M\right)$ с увеличением длины кадра M, то есть незначительно быстрее линейного, что обеспечивает масштабируемость алгоритма.

Блок ДВП-анализа (db4, $J = 4$, $K = 2$ канала):

После анализа Хургина – Яковлева каждый из $K = 2$ каналов содержит $M/K = 128$ отсчётов. ДВП выполняется независимо в каждом канале по быстрому алгоритму Малла: на каждом уровне j оба фильтра анализа — фильтр нижних частот (аппроксимация) и фильтр верхних частот (детали) — обрабатывают входную последовательность длиной $\left(M/K\right)/2^{j-1}$ отсчётов, выполняя L_h пар арифметических операций (1 умножение + 1 сложение) умножения-накопления на каждый отсчёт, где $L_h = 8$ — длина фильтра анализа для вейвлета db4. Суммирование по J уровням даёт число операций для одного канала:

$$C_{\text{ДВП}}^{(1)} = 2\frac{M}{K} L_h \left(2 - 2^{1-J}\right) = 2.128.8.1{,}875 = 3840 \text{ оп.}$$

Множитель $\left(2 - 2^{1-J}\right) = 1{,}875$ представляет собой сдвоенную сумму геометрической прогрессии со знаменателем $1/2$ по J уровням; удвоение обусловлено тем, что на каждом уровне оба выхода фильтрового банка — аппроксимирующие и детализирующие коэффициенты — подлежат энергетической пороговой обработке и, следовательно, должны быть вычислены в явном виде. Для $K = 2$ каналов:

$$C_{\text{ДВП}} = K.C_{\text{ДВП}}^{(1)} = 2.3840 = 7680 \text{ оп./кадр.}$$

Блок ОДВП на приёмной стороне: $C_{\text{ОДВП}} = C_{\text{ДВП}} = 7680$ оп./кадр. Поскольку суммарная стоимость по всем уровням ограничена удвоенным числом отсчётов первого уровня, зависимость от длины кадра линейна: $C_{\text{ДВП}} = \mathrm{O}\left(M\right)$.

Блок энергетического порогового сжатия:

Алгоритм порогового сжатия (шаг 1, формула 3.13) ранжирует вейвлет-коэффициенты на каждом из $(J+1)$ уровней каждого из K каналов. Доминирующей операцией является пирамидальная сортировка с гарантированной сложностью $\mathrm{O}(L_l \log L_l)$ на уровне l, где $L_l = (M/K)/2^l$ — число коэффициентов уровня l в одном канале. В модели вещественных операций каждый шаг сортировки включает одно сравнение и одну условную перестановку, что составляет 2 вещественные арифметические операции на шаг. Формула для полной стоимости:

$$C_{\text{пор}} = 2K \sum_{l=0}^{J} L_l \lceil \log_2 L_l \rceil.$$

При $K=2$, $M/K=128$ значения $L_l = \{64,32,16,8,8\}$ для $l = \{0,1,2,3,4\}$ соответственно. Вычисление суммы для одного канала осуществляется по формуле:

$$\sum_{l=0}^{4} L_l \log_2 L_l = 64.6 + 32.5 + 16.4 + 8.3 + 8.3 = 656 \text{ оп.},$$

$$C_{\text{пор}} = 2.2.656 = 2624 \text{ оп./кадр.}$$

Произведение $L_l \log_2 L_l$ суммируется по всем $(J+1)$ уровням, поэтому $C_{\text{пор}} = \mathrm{O}(M \log M)$. Среди трёх рассматриваемых блоков только блок порогового сжатия имеет надлинейную асимптотику, что необходимо учитывать при проектировании систем с большой длиной кадра.

Блок синтеза анализ Хургина – Яковлева ($K=2$, $P=1$):

На приёмной стороне K каналов по $M/K=128$ децимированных отсчётов. Синтезирующий фильтр каждого канала реализуется в частотной области как последовательность трёх операций: прямое БПФ длиной M/K для перехода в спектральную область, поэлементное умножение на предвычисленные коэффициенты $H_k[m]$, и обратное ОБПФ длиной M/K для возврата во временную область. Поскольку операция выполняется независимо для каждого из K каналов, полное число вещественных арифметических операций блока синтеза за один кадр:

$$C_{S,\times} = 2M \log_2 (M/K) \text{ (умножения)}, \qquad (3.46)$$

$$C_{S,+} = M\left[3\log_2\left(M/K\right) + (K-1)\right] \text{ (сложения).} \qquad (3.47)$$

При $M = 256$, $K = 2$: $C_{S,\times} = 3584$ (умножения), $C_{A,+} = 5632$ (сложения), получим полное число вещественных арифметических операций блока синтеза Хургина – Яковлева за один кадр:

$$C_S = C_{S,+} + C_{S,\times} = 5632 + 3584 = 9216 \text{ оп./кадр.}$$

Из структуры формулы (3.46), (3.47) следует, что при фиксированном K и увеличении длины кадра M стоимость блока синтеза растёт как $\mathrm{O}\left(M\log M\right)$, поскольку $\log_2\left(M/K\right) = \log_2 M - \log_2 K \sim \log_2 M$ при $M \to \infty$. Сравнение с блоком анализа: суммарный коэффициент при $M\log_2 M$ в формуле (3.46) и (3.47) равен $2+3=5$ (слагаемые $2M\log_2\left(M/K\right)$ и $3M\log_2\left(M/K\right)$, тогда как в формулах (3.44), (3.45) аналогичный коэффициент для блока анализа составляет $2(1+K)+3(1+K)=5(1+K)=15$ при $K=2$. Таким образом, при $M\to\infty$ блок анализа асимптотически в $(1+K)=3$ раза затратнее блока синтеза, что отражает принципиальную асимметрию системы: блок анализа на передатчике выполняет $(1+K)$ полноразмерных преобразований длиной M (одно прямое БПФ и K обратных ОБПФ до децимации), тогда как блок синтеза на приёмнике оперирует децимированными последовательностями длиной M/K. Данная асимметрия является нормой для многоскоростных систем анализа-синтеза [86].

Результаты анализа вычислительных сложности сведены в таблицу 3.3.

Из таблицы 3.3 следует, что блок анализа Хургина – Яковлева вносит 54,3 % от суммарной нагрузки — вследствие обработки полноразмерных последовательностей до децимации. Асимптотику системы в целом определяют блоки Хургина – Яковлева и пороговое сжатие с $\mathrm{O}\left(M\log M\right)$; блоки ДВП/ОДВП имеют линейную O(M), что обеспечивает масштабируемость при увеличении M.

Для перехода к оценке нагрузки в реальном времени введём $f_{\text{кадр}}$ — частоту следования кадров (кадр/с), определяемую как отношение частоты дискретизации к длине кадра:

$$f_{\text{кадр}} = \frac{8000}{M} = \frac{8000}{256} = 31{,}25 \text{ кадр/с.}$$

Таблица 3.3. Вычислительная сложность основных блоков интегрированной системы

Функциональный блок	Оп./кадр	Доля, %	O(·)
Анализ Хургина – Яковлева	32256	54,3	$\mathrm{O}(M \log M)$
ДВП-анализ	7680	12,9	$\mathrm{O}(M)$
Пороговое сжатие	2624	4,4	$\mathrm{O}(M \log M)$
ОДВП-синтез	7680	12,9	$\mathrm{O}(M)$
Синтез Хургина – Яковлева	9216	15,5	$\mathrm{O}(M \log M)$
ИТОГО $C_{\text{РВ}}$	59456	100	$\mathrm{O}(M \log M)$

Требуемая вычислительная производительность для обработки одного речевого канала в режиме реального времени:

$$\Pi_{\text{РВ}} = C_{\text{РВ}} f_{\text{кадр}} = 59456.31{,}25/10^6 \approx 1{,}86 \text{ MIPS.} \quad (3.48)$$

Консервативная оценка с коэффициентом архитектурного запаса $k = 1{,}8$ (учитывает простои конвейера, промахи кэша и служебные инструкции, типичный диапазон $k = 1{,}4...1{,}8$ для ДСП- архитектур [86]):

$$\Pi_{\text{КОНС}} = k\Pi_{\text{РВ}} = 1{,}8.1{,}86 \approx 3{,}34 \text{ MIPS.} \quad (3.49)$$

Для практических расчётов в разделе 3.5.2 используется консервативная оценка $\Pi_{\text{КОНС}} = 3{,}34$ MIPS, обеспечивающая запас надёжности при реализации на реальной платформе.

3.5.2 Требования к аппаратным платформам и рекомендации по выбору элементной базы

Из результатов раздела 3.5.1 следуют количественные требования к аппаратной платформе для реализации интегрированной системы в режиме реального времени. Платформа должна обеспечивать вычислительную

производительность не менее $\Pi_{\text{КОНС}} = 3{,}34$ MIPS на один речевой канал. Максимальное число каналов, обслуживаемых одной платформой, и относительная нагрузка определяются соотношениями:

$$N_{\text{кан}} = \left[\frac{\Pi_{\text{платф}}}{\Pi_{\text{КОНС}}} \right], \ \eta = \frac{\Pi_{\text{КОНС}}}{\Pi_{\text{платф}}}.100\%, \tag{3.50}$$

где $\Pi_{\text{платф}}$ — паспортная производительность платформы в MIPS. На основании данного критерия рассматриваются четыре класса платформ, охватывающие основные области применения встраиваемых радиотехнических систем. Характеристики платформ приняты по техническим описаниям производителей [101–104] и сведены в таблицу 3.4.

Таблица 3.4. Сравнительная характеристика аппаратных платформ для реализации интегрированной системы ($K = 2$, $P = 1$, $M = 256$)

Параметр	TMS320C6748 [101]	ADSP-21569 [102]	Zynq-7020 PS [103]	Cortex-M55 [104]
Тактовая частота (МГц)	456	1000 (2 ядра)	667 (ARM A9)	800
Производительность $\Pi_{\text{платф}}$ (MIPS)	456	2000	667	800
Относительная нагрузка η,%	0,73	0,17	0,50	0,42
Максимальное число каналов $N_{\text{кан}}$	~136	~598	~199	~239
Потребляемая мощность (мВт)	≤ 400	≤ 2000	≤ 1300	< 10

Малая относительная нагрузка $\eta \leq 0{,}73\%$ для всех рассматриваемых платформ подтверждает, что интегрированная система не является вычислительно лимитирующим блоком ни в одном из рассматриваемых классов оборудования, а

запас производительности допускает параллельную реализацию дополнительных функций связи без аппаратного расширения.

Цифровой сигнальный процессор TMS320C6748 при нагрузке $\eta = 0,73\%$ и потреблении не более 400 мВт является предпочтительной платформой для носимых тактических терминалов; обеспечивая 136 гарантированных речевых каналов.

Цифровой сигнальный процессор ADSP-21569 обеспечивает обслуживание порядка 598 каналов при $\eta = 0,17\%$, что делает его оптимальным для многоканальных шлюзов и базовых станций.

Система на кристалле Zynq-7020, объединяющая процессорное ядро и программируемую логическую интегральную схему, обеспечивает около 199 каналов на процессорном ядре. Возможность реализации тракта обработки в программируемой логике позволяет дополнительно снизить нагрузку на процессор и обеспечить детерминированную задержку обработки.

Микроконтроллер ARM Cortex-M55 с потреблением менее 10 мВт предназначен для малогабаритных терминалов и беспилотных летательных аппаратов. При $\eta = 0,42\%$ он обеспечивает 239 каналов, что подтверждает пригодность платформы для применения в условиях жёстких ограничений по массогабаритным характеристикам и энергопотреблению.

Таким образом, требуемая производительность $\Pi_{\text{КОНС}} = 3,34$ MIPS на один речевой канал не накладывает существенных ограничений ни на один из рассмотренных классов аппаратных платформ. Максимальная консервативная нагрузка 0,73 % (TMS320C6748) свидетельствует о том, что интегрированная система пригодна для реализации на широком спектре встраиваемых платформ без специальных мер по оптимизации вычислительного процесса.

3.6 Выводы

В третьей главе разработаны метод вейвлет-сжатия РС с покадровой обработкой, метод адаптивной оптимизации энергетических порогов на основе алгоритма роя частиц PSO, интегрированная система совместной обработки на основе представления Хургина – Яковлева и вейвлет-преобразования, а также

обоснованы предложения по реализации разработанных алгоритмов на встраиваемых платформах. По результатам выполненных исследований сформулированы следующие выводы.

1. Разработан метод покадрового ДВП с энергетическим критерием порогового сжатия, принципиально отличающийся от классических подходов раздельным управлением степенью сжатия в каждой частотной полосе посредством вектора уровневых порогов $\alpha \in [0,1]^{J+1}$. Экспериментальное сравнение 17 вейвлет-базисов четырёх семейств установило, что вейвлет Добеши db4 при $J = 4$ уровнях разложения обеспечивает наилучший интегральный критерий эффективности. Оптимум $J = 4$ является инвариантным к типу вейвлета и физически обусловлен спектральной структурой речевого сигнала.

2. Сформулирована задача многокритериальной оптимизации вектора энергетических порогов как задача нелинейного программирования с негладкой многоэкстремальной целевой функцией, и разработан алгоритм её решения на основе метода роя частиц PSO. Алгоритм сошёлся к оптимальному решению на 48-й итерации из 100 максимально допустимых, что подтверждает эффективность критерия ранней остановки. Достигнутый коэффициент сжатия $\rho = 87\%$ при $Q_1 = 9{,}5$ дБ и $Q_2 = 3{,}87$ балла превосходит метод Донохо – Джонстона на 14,7 %, SUREShrink — на 7,5 %, фиксированный энергетический порог — на 5,8 %.

3. Разработана и исследована интегрированная система первичного кодирования речи на основе представления Хургина – Яковлева ($K = 2$, $P = 1$), ДВП (db4, $J = 4$) и оптимизации энергетических порогов методом роя частиц PSO в пространстве размерности $D = 10$. Достигнутая степень разреженности $\rho \approx 93\%$ превышает результат системы ДВП без представления Хургина – Яковлева на 6% при неизменных ограничениях на качество. Экспериментальное исследование при низкоскоростном кодировании (LBRAMR, C = 1,0…5,0 кбит/с, $P_{ош} = 0\ldots5\%$) показало, что интегрированная система обеспечивает выигрыш по метрике ViSQOL до 0,9 балла по шкале MOS по сравнению с системой на основе теоремы В.А. Котельникова без увеличения скорости передачи данных.

4. Выполнен анализ вычислительной сложности алгоритмов интегрированной системы. Разработаны рекомендации по реализации системы на платформах TMS320C6748, ADSP-21569, Zynq-7020 и Cortex-M55, подтвердившие полную аппаратную реализуемость системы без компромисса по качеству обработки.

ЗАКЛЮЧЕНИЕ

В диссертационной работе решены поставленные задачи и получены следующие основные результаты:

1. Разработана математическая модель системы первичного кодирования РС на основе представления Хургина – Яковлева. Установлена эквивалентность данной системы K-канальному банку фильтров анализа-синтеза с максимальной децимацией, анализирующие фильтры которого определяются оператором конечных разностей порядка k. Доказано, что биномиальная матрица полифазных компонент обладает свойством инволютивности, что обеспечивает выполнение условия полного восстановления дискретного сигнала тождественно при любом числе каналов K.

2. Разработан и исследован метод синтеза АЧХ и ФЧХ синтезирующих фильтров представления Хургина – Яковлева по критерию минимума среднеквадратического отклонения для двухканальной ($K = 2$), трёхканальной ($K = 3$) и четырёхканальной ($K = 4$) конфигураций системы. Для двухканальной системы синтезирующие фильтры определяются аналитически в замкнутой форме, что гарантирует точное восстановление непрерывного РС. Для конфигураций $K = 3$ и $K = 4$ применяется кусочно-линейная аппроксимация. Предложенная модификация синтезирующих фильтров для двухканальной системы обеспечивает точность восстановления SNR = 85,38 дБ при MSE = $7{,}6 \cdot 10^{-12}$ и ViSQOL = 4,73 балла при одноэтапном разложении ($P = 1$).

3. Использование многоканальной и многоуровневой структуры позволяет значительно снизить вычислительные затраты по сравнению с классическими алгоритмами обработки, основанными на теореме дискретизации В.А. Котельникова. Для системы с двумя каналами ($K = 2$) и числом уровней разложений $P = 1\ldots3$ относительное сокращение вычислительных затрат достигает диапазона от 1,143 до 1,75 раз для кадров анализа размером $M = 128$ и $M = 256$, соответственно.

4. Модификация синтезирующих фильтров для представления Хургина – Яковлева в двухканальной структуре с двух- или трёхуровневым

разложением обеспечивает точность восстановления не менее 45 дБ, что свидетельствует о высоком качестве восстановления речи. Для трёх- и четырёхканальных структур точность восстановления составляет не менее 14 дБ, что является достаточным для систем передачи речи с низкой скоростью передачи. Использование структур с числом каналов более четырёх нецелесообразно, поскольку приводит к снижению ОСШ и не даёт выигрыша в вычислительной эффективности.

5. Обоснована и экспериментально подтверждена эффективность алгоритмов первичного кодирования РС на основе представления Хургина – Яковлева в сравнении с классической системой дискретизации В.А. Котельникова. В условиях воздействия аддитивного белого гауссовского шума аналитически рассчитан выигрыш в ОСШ, составляющий: 2,27 дБ ($K = 2$), 6,20 дБ ($K = 3$) и 8,68 дБ ($K = 4$) для одноэтапного разложения. Экспериментальное исследование на выборке из 100 речевых записей показало, что алгоритмы на основе представления Хургина – Яковлева обеспечивают выигрыш в качестве восстановленной речи по шкале MOS до 2,2 балла (для кодека G.723.1 при $K = 4$, $P = 3$, $P_{ош} = 5\%$) без увеличения скорости передачи данных.

6. Предложен метод адаптивной оптимизации энергетических порогов вейвлет-преобразования с использованием алгоритма роя частиц (PSO). Он автоматически выбирает доли сохраняемой энергии на каждом уровне, максимизируя степень сжатия при заданном качестве восстановления. Достигнута степень сжатия 87% при SegSNR = 9,5 дБ и ViSQOL = 3,9 балла, что на 7…14% превосходит известные методы по интегральному критерию эффективности.

7. Интегрированная модель первичного кодирования речи на основе представления Хургина – Яковлева, вейвлет-преобразования и оптимизации энергетических порогов методом роя частиц обеспечивает выигрыш в качестве до 0,9 балла по шкале MOS в канале с ошибками 1…5 %, что подтверждает эффективность решения для низкоскоростного диапазона 1…3 кбит/с.

СПИСОК ЛИТЕРАТУРЫ

1. Котельников, В. А. О пропускной способности «эфира» и проволоки в электросвязи: переиздание доклада 1933 г. // Успехи физических наук. — 2006. — Т. 176, № 7. — С. 762–770. — DOI: 10.3367/UFNr.0176.200607f.0762.

2. Shannon, C. E. Communication in the Presence of Noise // Proceedings of the IRE. — 1949. — Vol. 37, No. 1. — P. 10–21. — DOI: 10.1109/JRPROC.1949.232969.

3. Shannon, C. E. A Mathematical Theory of Communication // Bell System Technical Journal. — 1948. — Vol. 27, No. 3. — P. 379–423. — DOI: 10.1002/j.1538-7305.1948.tb01338.x.

4. Whittaker, E. T. On the Functions Which Are Represented by the Expansions of the Interpolation Theory // Proceedings of the Royal Society of Edinburgh. — 1915. — Vol. 35. — P. 181–194. — DOI: 10.1017/S0370164600017806.

5. Whittaker, J. M. Interpolatory Function Theory. — Cambridge: Cambridge University Press, 1935. — 144 p.

6. Nyquist, H. Certain Topics in Telegraph Transmission Theory // Transactions of the American Institute of Electrical Engineers. — 1928. — Vol. 47, No. 2. — P. 617–644. — DOI: 10.1109/T-AIEE.1928.5055024.

7. Jerri, A. J. The Shannon Sampling Theorem — Its Various Extensions and Applications: A Tutorial Review // Proceedings of the IEEE. — 1977. — Vol. 65, No. 11. — P. 1565–1596. — DOI: 10.1109/PROC.1977.10771.

8. Оппенгейм, А. В. Цифровая обработка сигналов / А. В. Оппенгейм, Р. В. Шафер; пер. с англ. С. К. Ситникова. — М.: Техносфера, 2007. — 856 с. — ISBN 978-5-94836-135-2.

9. Unser, M. Sampling — 50 Years After Shannon // Proceedings of the IEEE. — 2000. — Vol. 88, No. 4. — P. 569–587. — DOI: 10.1109/5.843002.

10. Marks, R. J. Introduction to Shannon Sampling and Interpolation Theory. — New York: Springer-Verlag, 1991. — 336 p. — ISBN 978-0-387-97391-0.

11. Хургин, Я. И. Финитные функции в физике и технике / Я. И. Хургин, В. П. Яковлев. — М.: Наука, 1971. — 408 с.

12. Хургин, Я. И. Методы теории целых функций в радиофизике, теории связи

и оптике / Я. И. Хургин, В. П. Яковлев. — М.: ГИФМЛ, 1962. — 220 с.

13. Яковлев, В. П. Периодические пространства финитных функций. — М.: Московский энергетический институт, 1990. — 116 с.

14. Хургин, Я. И. Прогресс в Советском Союзе в области теории финитных функций и её применений в физике и технике / Я. И. Хургин, В. П. Яковлев // ТИЭР. — 1977. — Т. 65, № 7. — С. 16–45.

15. Papoulis, A. Generalized Sampling Expansion // IEEE Transactions on Circuits and Systems. — 1977. — Vol. 24, No. 11. — P. 652–654. — DOI: 10.1109/TCS.1977.1084284.

16. Vetterli, M. Foundations of Signal Processing / M. Vetterli, J. Kovačević, V. K. Goyal. — Cambridge: Cambridge University Press, 2014. — 716 p. — ISBN 978-1-107-03860-8.

17. Dragotti, P. L. Sampling Moments and Reconstructing Signals of Finite Rate of Innovation: Shannon Meets Strang–Fix / P. L. Dragotti, M. Vetterli, T. Blu // IEEE Transactions on Signal Processing. — 2007. — Vol. 55, No. 5. — P. 1741–1757. — DOI: 10.1109/TSP.2006.890907.

18. Vetterli, M. Sampling Signals with Finite Rate of Innovation / M. Vetterli, P. Marziliano, T. Blu // IEEE Transactions on Signal Processing. — 2002. — Vol. 50, No. 6. — P. 1417–1428. — DOI: 10.1109/TSP.2002.1003065.

19. Blu, T. Sparse Sampling of Signal Innovations / T. Blu, P. L. Dragotti, M. Vetterli [et al.] // IEEE Signal Processing Magazine. — 2008. — Vol. 25, No. 2. — P. 31–40. — DOI: 10.1109/MSP.2007.914998.

20. Eldar, Y. C. Sampling Theory: Beyond Bandlimited Systems. — Cambridge: Cambridge University Press, 2015. — 820 p. — ISBN 978-1-107-00356-9.

21. Vaidyanathan, P. P. Multirate Systems and Filter Banks. — Englewood Cliffs: Prentice Hall, 1993. — 911 p. — ISBN 978-0-13-605718-4.

22. Crochiere, R. E. Multirate Digital Signal Processing / R. E. Crochiere, L. R. Rabiner. — Englewood Cliffs: Prentice-Hall, 1983. — 411 p. — ISBN 978-0-13-605162-6.

23. Витязев, В. В. Многоскоростная обработка сигналов. — М.: Горячая линия–Телеком, 2017. — 336 с. — ISBN 978-5-9912-0608-5.

24. Витязев, В. В. Цифровая частотная селекция сигналов. — М.: Радио и связь, 1993. — 240 с. — ISBN 5-256-01029-8.

25. Солонина, А. И. Алгоритмы и процессоры цифровой обработки сигналов / А. И. Солонина, Д. А. Улахович, Л. А. Яковлев. — СПб.: БХВ-Петербург, 2001. — 464 с. — ISBN 5-94157-097-5.

26. Strang, G. Wavelets and Filter Banks / G. Strang, T. Nguyen. — 2nd ed. — Wellesley: Wellesley-Cambridge Press, 1996. — 520 p. — ISBN 978-0-9614088-7-2.

27. Bellanger, M. G. Digital Filtering by Polyphase Network: Application to Sample Rate Alteration and Filter Banks / M. G. Bellanger, J. L. Daguet, G. P. Lepagnol // IEEE Transactions on Acoustics, Speech, and Signal Processing. — 1976. — Vol. 24, No. 2. — P. 109–114. — DOI: 10.1109/TASSP.1976.1162788.

28. Vetterli, M. Wavelets and Filter Banks: Theory and Design / M. Vetterli, C. Herley // IEEE Transactions on Signal Processing. — 1992. — Vol. 40, No. 9. — P. 2207–2232. — DOI: 10.1109/78.157221.

29. Mallat, S. G. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 1989. — Vol. 11, No. 7. — P. 674–693. — DOI: 10.1109/34.192463.

30. Малла, С. Вейвлеты в обработке сигналов / пер. с англ. — М.: Мир, 2005. — 671 с. — ISBN 5-03-003576-8.

31. Добеши, И. Десять лекций по вейвлетам / пер. с англ. — Ижевск: НИЦ «Регулярная и хаотическая динамика», 2001. — 464 с. — ISBN 5-93972-012-6.

32. Daubechies, I. Orthonormal Bases of Compactly Supported Wavelets // Communications on Pure and Applied Mathematics. — 1988. — Vol. 41, No. 7. — P. 909–996. — DOI: 10.1002/cpa.3160410705.

33. Coifman, R. R. Entropy-Based Algorithms for Best Basis Selection / R. R. Coifman, M. V. Wickerhauser // IEEE Transactions on Information Theory. — 1992. — Vol. 38, No. 2. — P. 713–718. — DOI: 10.1109/18.119732.

34. Cohen, A. Biorthogonal Bases of Compactly Supported Wavelets / A. Cohen, I. Daubechies, J.-C. Feauveau // Communications on Pure and Applied Mathematics. — 1992. — Vol. 45, No. 5. — P. 485–560. — DOI: 10.1002/cpa.3160450502.

35. Unser, M. Approximation Power of Biorthogonal Wavelet Expansions // IEEE Transactions on Signal Processing. — 1996. — Vol. 44, No. 3. — P. 519–527. — DOI: 10.1109/78.489025.

36. Sweldens, W. The Lifting Scheme: A Construction of Second Generation Wavelets // SIAM Journal on Mathematical Analysis. — 1998. — Vol. 29, No. 2. — P. 511–546. — DOI: 10.1137/S0036141095289051.

37. Vetterli, M. Wavelets and Subband Coding / M. Vetterli, J. Kovačević. — Englewood Cliffs: Prentice Hall, 1995. — 488 p. — ISBN 978-0-13-097079-5.

38. Новиков, И. Я. Теория всплесков / И. Я. Новиков, В. Ю. Протасов, М. А. Скопина. — М.: ФИЗМАТЛИТ, 2005. — 616 с. — ISBN 5-9221-0642-4.

39. Новиков, И. Я. Основы теории всплесков / И. Я. Новиков, С. Б. Стечкин // Успехи математических наук. — 1998. — Т. 53, № 6 (324). — С. 53–128. — DOI: 10.4213/rm91.

40. Скопина, М. А. Многомерные ортогональные базисы всплесков // Алгебра и анализ. — 1996. — Т. 8, № 4. — С. 167–192.

41. Лебедева, Е. А. О минимальном свойстве неопределённости в базисах всплесков / Е. А. Лебедева, И. Я. Новиков // Сибирский математический журнал. — 2010. — Т. 51, № 3. — С. 619–627. — DOI: 10.1007/s11202-010-0060-0.

42. Лебедева, Е. А. Минимизация неопределённости в многовейвлетном анализе // Сибирский математический журнал. — 2012. — Т. 53, № 1. — С. 144–154. — DOI: 10.1134/S0037446612010107.

43. Фарков, Ю. А. Ортогональные вейвлеты с компактным носителем на локально компактных абелевых группах // Известия РАН. Серия математическая. — 2005. — Т. 69, № 3. — С. 193–220. — DOI: 10.4213/im641.

44. Астафьева, Н. М. Вейвлет-анализ: основы теории и примеры применения // Успехи физических наук. — 1996. — Т. 166, № 11. — С. 1145–1170. — DOI: 10.3367/UFNr.0166.199611a.1145.

45. Дремин, И. М. Вейвлеты и их использование / И. М. Дремин, О. В. Иванов, В. А. Нечитайло // Успехи физических наук. — 2001. — Т. 171, № 5. — С. 465–501. — DOI: 10.3367/UFNr.0171.200105b.0465.

46. Воробьев, В. И. Теория и практика вейвлет-преобразования / В. И. Воробьев, В. Г. Грибунин. — СПб.: ВУС, 1999. — 208 с.

47. Дьяконов, В. П. Вейвлеты: от теории к практике. — 2-е изд. — М.: СОЛОН-Р, 2002. — 448 с. — ISBN 5-98003-011-0.

48. Новиков, Л. В. Основы вейвлет-анализа сигналов: учеб. пособие. — СПб.: ИАнП РАН, 1999. — 152 с.

49. Смоленцев, Н. К. Основы теории вейвлетов. Вейвлеты в MATLAB. — М.: ДМК Пресс, 2008. — 448 с. — ISBN 978-5-94074-450-7.

50. Чуи, К. Введение в вейвлеты / пер. с англ. — М.: Мир, 2001. — 412 с. — ISBN 5-03-003441-9.

51. Donoho, D. L. De-Noising by Soft-Thresholding // IEEE Transactions on Information Theory. — 1995. — Vol. 41, No. 3. — P. 613–627. — DOI: 10.1109/18.382009.

52. Donoho, D. L. Ideal Spatial Adaptation by Wavelet Shrinkage / D. L. Donoho, I. M. Johnstone // Biometrika. — 1994. — Vol. 81, No. 3. — P. 425–455. — DOI: 10.1093/biomet/81.3.425.

53. Donoho, D. L. Adapting to Unknown Smoothness via Wavelet Shrinkage / D. L. Donoho, I. M. Johnstone // Journal of the American Statistical Association. — 1995. — Vol. 90, No. 432. — P. 1200–1224. — DOI: 10.1080/01621459.1995.10476626.

54. Coifman, R. R. Translation-Invariant De-noising / R. R. Coifman, D. L. Donoho // Wavelets and Statistics: Lecture Notes in Statistics. — New York: Springer, 1995. — Vol. 103. — P. 125–150. — DOI: 10.1007/978-1-4612-2544-7_9.

55. Mallat, S. G. Multiresolution Approximations and Wavelet Orthonormal Bases of $L^2(\mathbb{R})$ // Transactions of the American Mathematical Society. — 1989. — Vol. 315, No. 1. — P. 69–87. — DOI: 10.2307/2001373.

56. Kingsbury, N. Complex Wavelets for Shift Invariant Analysis and Filtering of Signals // Applied and Computational Harmonic Analysis. — 2001. — Vol. 10, No. 3. — P. 234–253. — DOI: 10.1006/acha.2000.0343.

57. Kennedy, J. Particle Swarm Optimization / J. Kennedy, R. Eberhart // Proceedings of the IEEE International Conference on Neural Networks (ICNN'95). —

Piscataway: IEEE, 1995. — Vol. 4. — P. 1942–1948. — DOI: 10.1109/ICNN.1995.488968.

58. Eberhart, R. C. A New Optimizer Using Particle Swarm Theory / R. C. Eberhart, J. Kennedy // Proceedings of the 6th International Symposium on Micro Machine and Human Science (MHS'95). — Piscataway: IEEE, 1995. — P. 39–43. — DOI: 10.1109/MHS.1995.494215.

59. Shi, Y. A Modified Particle Swarm Optimizer / Y. Shi, R. C. Eberhart // Proceedings of the IEEE World Congress on Evolutionary Computation (CEC'98). — Piscataway: IEEE, 1998. — P. 69–73. — DOI: 10.1109/ICEC.1998.699146.

60. Shi, Y. Empirical Study of Particle Swarm Optimization / Y. Shi, R. C. Eberhart // Proceedings of the 1999 Congress on Evolutionary Computation (CEC'99). — Piscataway: IEEE, 1999. — Vol. 3. — P. 1945–1950. — DOI: 10.1109/CEC.1999.785511.

61. Clerc, M. The Particle Swarm — Explosion, Stability, and Convergence in a Multidimensional Complex Space / M. Clerc, J. Kennedy // IEEE Transactions on Evolutionary Computation. — 2002. — Vol. 6, No. 1. — P. 58–73. — DOI: 10.1109/4235.985692.

62. Bratton, D. Defining a Standard for Particle Swarm Optimization / D. Bratton, J. Kennedy // Proceedings of the 2007 IEEE Swarm Intelligence Symposium (SIS'07). — Piscataway: IEEE, 2007. — P. 120–127. — DOI: 10.1109/SIS.2007.368035.

63. Poli, R. Particle Swarm Optimization: An Overview / R. Poli, J. Kennedy, T. Blackwell // Swarm Intelligence. — 2007. — Vol. 1, No. 1. — P. 33–57. — DOI: 10.1007/s11721-007-0002-0.

64. Hu, X. Multiobjective Optimization Using Dynamic Neighbourhood Particle Swarm Optimization / X. Hu, R. C. Eberhart // Proceedings of the 2002 Congress on Evolutionary Computation (CEC'02). — Piscataway: IEEE, 2002. — Vol. 2. — P. 1677–1681. — DOI: 10.1109/CEC.2002.1004494.

65. Engelbrecht, A. P. Fundamentals of Computational Swarm Intelligence. — Chichester: John Wiley & Sons, 2006. — 672 p. — ISBN 978-0-470-09191-3.

66. Van den Bergh, F. A Cooperative Approach to Particle Swarm Optimization /

F. van den Bergh, A. P. Engelbrecht // IEEE Transactions on Evolutionary Computation. — 2004. — Vol. 8, No. 3. — P. 225–239. — DOI: 10.1109/TEVC.2004.826069.

67. Карпенко, А. П. Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновлённые природой: учеб. пособие. — 2-е изд. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2017. — 446 с. — ISBN 978-5-7038-4634-6.

68. Гладков, Л. А. Биоинспирированные методы в оптимизации / Л. А. Гладков, В. В. Курейчик, В. М. Курейчик [и др.]. — М.: Физматлит, 2009. — 384 с. — ISBN 978-5-9221-1086-6.

69. Пантелеев, А. В. Методы глобальной оптимизации: метаэвристические стратегии и алгоритмы / А. В. Пантелеев, Д. В. Метлицкая, Е. А. Алёшина. — М.: Вузовская книга, 2013. — 244 с. — ISBN 978-5-9502-0545-7.

70. Mezura-Montes, E. Constraint-Handling in Nature-Inspired Numerical Optimization: Past, Present and Future / E. Mezura-Montes, C. A. Coello Coello // Swarm and Evolutionary Computation. — 2011. — Vol. 1, No. 4. — P. 173–194. — DOI: 10.1016/j.swevo.2011.10.001.

71. French, N. R. Factors Governing the Intelligibility of Speech Sounds / N. R. French, J. C. Steinberg // Journal of the Acoustical Society of America. — 1947. — Vol. 19, No. 1. — P. 90–119. — DOI: 10.1121/1.1916407.

72. ITU-T Recommendation P.800. Methods for Subjective Determination of Transmission Quality. — Geneva: ITU-T, 1996. — 38 p.

73. ГОСТ Р 50840-95. Передача речи по трактам и каналам связи. Методы оценки качества, разборчивости и узнаваемости. — М.: Издательство стандартов, 1996. — 234 с.

74. Chinen, M. ViSQOL v3: An Open Source Production Ready Objective Speech and Audio Metric / M. Chinen, J. Skoglund, A. Luebs [et al.] // Proceedings of Interspeech 2020. — Shanghai, 2020. — P. 4080–4084. — DOI: 10.21437/Interspeech.2020-1298.

75. ITU-T Recommendation G.114. One-Way Transmission Time. — Geneva: ITU-T, 2003. — 12 p.

76. ITU-T Recommendation G.711. Pulse Code Modulation (PCM) of Voice Frequencies. — Geneva: ITU-T, 1988. — 12 p.

77. ITU-T Recommendation G.726. 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM). — Geneva: ITU-T, 1990. — 68 p.

78. ETSI EN 300 961. Digital Cellular Telecommunications System (Phase 2+): Full Rate Speech; Transcoding (GSM 06.10). — Sophia Antipolis: ETSI, 2000. — 46 p.

79. ITU-T Recommendation G.729. Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP). — Geneva: ITU-T, 1996. — 75 p.

80. ITU-T Recommendation G.723.1. Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s. — Geneva: ITU-T, 1996. — 49 p.

81. McCree, A. V. A Mixed Excitation LPC Vocoder Model for Low Bit Rate Speech Coding / A. V. McCree, T. P. Barnwell III // IEEE Transactions on Speech and Audio Processing. — 1995. — Vol. 3, No. 4. — P. 242–250. — DOI: 10.1109/89.397090.

82. 3GPP TS 26.090 v7.0.0. Adaptive Multi-Rate (AMR) Speech Codec; Transcoding Functions. — Sophia Antipolis: 3GPP, 2007. — 80 p.

83. Flanagan, J. L. Speech Analysis, Synthesis and Perception. — 2nd ed. — Berlin: Springer-Verlag, 1972. — 444 p. — ISBN 978-3-540-05618-5.

84. Rabiner, L. R. Theory and Applications of Digital Speech Processing / L. R. Rabiner, R. W. Schafer. — Upper Saddle River: Prentice Hall, 2011. — 1056 p. — ISBN 978-0-13-601638-8.

85. Рабинер, Л. Цифровая обработка речевых сигналов / Л. Рабинер, Р. Шафер; пер. с англ. под ред. М. В. Назарова, Ю. Н. Прохорова. — М.: Радио и связь, 1981. — 496 с.

86. Quatieri, T. F. Discrete-Time Speech Signal Processing: Principles and Practice. — Upper Saddle River: Prentice Hall, 2002. — 712 p. — ISBN 978-0-13-228438-0.

87. Jayant, N. S. Digital Coding of Waveforms: Principles and Applications to Speech and Video / N. S. Jayant, P. Noll. — Englewood Cliffs: Prentice Hall, 1984. — 688 p. — ISBN 978-0-13-211913-0.

88. Loizou, P. C. Speech Enhancement: Theory and Practice. — 2nd ed. — Boca Raton: CRC Press, 2013. — 900 p. — ISBN 978-1-4665-0421-8.

89. Atal, B. S. Predictive Coding of Speech Signals and Subjective Error Criteria /

B. S. Atal, M. R. Schroeder // IEEE Transactions on Acoustics, Speech, and Signal Processing. — 1979. — Vol. 27, No. 3. — P. 247–254. — DOI: 10.1109/TASSP.1979.1163237.

90. Schroeder, M. R. Code-Excited Linear Prediction (CELP): High-Quality Speech at Very Low Bit Rates / M. R. Schroeder, B. S. Atal // Proceedings of ICASSP-85. — Tampa, 1985. — P. 937–940. — DOI: 10.1109/ICASSP.1985.1168147.

91. Шелухин, О. И. Цифровая обработка и передача речи / О. И. Шелухин, А. М. Тенякшев, А. В. Осин. — М.: Радио и связь, 2001. — 228 с. — ISBN 5-256-01558-3.

92. Прокис, Дж. Цифровая связь / Пер. с англ. под ред. Д. Д. Кловского. — М.: Радио и связь, 2000. — 800 с. — ISBN 5-256-01434-X.

93. Гоноровский, И. С. Радиотехнические цепи и сигналы : учебник для вузов. — 5-е изд. — М.: Радио и связь, 1994. — 480 с. — ISBN 5-256-01060-3.

94. Гольденберг, Л. М. Цифровая обработка сигналов: учеб. пособие / Л. М. Гольденберг, Б. Д. Матюшкин, М. Н. Поляк. — М.: Радио и связь, 1990. — 256 с. — ISBN 5-256-00477-2.

95. Баскаков, С. И. Радиотехнические цепи и сигналы: учебник для вузов. — 4-е изд. — М.: Высшая школа, 2003. — 462 с. — ISBN 5-06-003824-1.

96. Сергиенко, А. Б. Цифровая обработка сигналов: учебник. — 2-е изд. — СПб.: Питер, 2006. — 752 с. — ISBN 978-5-469-01025-0.

97. Лайонс, Р. Понимание цифровой обработки сигналов / Пер. с англ. — 2-е изд. — М.: Бином-Пресс, 2006. — 656 с. — ISBN 5-9518-0038-7.

98. Витязев, С. В. Цифровые процессоры обработки сигналов: учеб. пособие. — М.: Горячая линия–Телеком, 2017. — 184 с. — ISBN 978-5-9912-0640-5.

99. Хайкин, С. Нейронные сети: полный курс / Пер. с англ. — 2-е изд. — М.: Вильямс, 2006. — 1104 с. — ISBN 5-8459-0890-6.

100. Percival, D. B. Wavelet Methods for Time Series Analysis / D. B. Percival, A. T. Walden. — Cambridge: Cambridge University Press, 2000. — 620 p. — ISBN 978-0-521-68508-5.

101. Texas Instruments. TMS320C6748 Fixed/Floating-Point DSP: Technical Reference Manual SPRUGK5A. — Dallas: Texas Instruments, 2010. — 830 p. — URL:

https://www.ti.com/lit/ug/sprugk5a/sprugk5a.pdf (дата обращения: 10.01.2025).

102. Analog Devices Inc. ADSP-21569 SHARC+ Processor Hardware Reference (Rev. 0.3). — Norwood: Analog Devices, 2019. — URL: https://www.analog.com/media/en/dsp-documentation/processor-manuals/ADSP-21569_Hardware_Reference.pdf (дата обращения: 10.01.2025).

103. AMD (Xilinx). Zynq-7000 SoC Technical Reference Manual UG585 (v1.13). — San Jose: AMD, 2023. — URL: https://docs.xilinx.com/r/en-US/ug585-zynq-7000-TRM (дата обращения: 10.01.2025).

104. Arm Limited. Cortex-M55 Processor Technical Reference Manual r1p0. — Cambridge: Arm Limited, 2022. — URL: https://developer.arm.com/documentation/101273/r1p0 (дата обращения: 10.01.2025).

105. Дмитриев, В. Т. Применение трехканальной модификации алгоритма Хургина – Яковлева в алгоритмах первичного кодирования речевых сигналов / В. Т. Дмитриев, Ву Хоанг Шон // Вестник Рязанского государственного радиотехнического университета. — 2024. — № 88. — С. 3–14.

106. Дмитриев, В. Т. Разработка многоканальной системы на основе модификации алгоритма Хургина – Яковлева при использовании отсчетов сигнала и его трех первых производных / В. Т. Дмитриев, Ву Хоанг Шон // Вестник Рязанского государственного радиотехнического университета. — 2025. — № 91. — С. 24–35.

107. Дмитриев, В. Т. Разработка и исследование первичных кодеков речи при многоэтапном разложении на основе модификации алгоритма Хургина – Яковлева / В. Т. Дмитриев, Ву Хоанг Шон // Цифровая обработка сигналов. — 2025. — № 3. — С. 76–85.

108. Дмитриев, В. Т. Оптимизация энергетических порогов вейвлет-преобразования для сжатия речевых сигналов на основе алгоритма роя частиц / В. Т. Дмитриев, Ву Хоанг Шон // Вестник Рязанского государственного радиотехнического университета. — 2026. — № 95. — С. 3–15.

109. Дмитриев, В. Т. Программа моделирования модификации алгоритма Хургина – Яковлева при трехканальной реализации: свидетельство о

государственной регистрации программы для ЭВМ № 2024619863 / В. Т. Дмитриев, Ву Хоанг Шон; правообладатель ФГБОУ ВО «Рязанский государственный радиотехнический университет имени В. Ф. Уткина». — Зарегистрировано в Реестре программ для ЭВМ 27.04.2024.

110. Дмитриев, В. Т. Программа для исследования синтезирующих фильтров Хургина – Яковлева: свидетельство о государственной регистрации программы для ЭВМ № 2025618510 / В. Т. Дмитриев, Ву Хоанг Шон; правообладатель ФГБОУ ВО «Рязанский государственный радиотехнический университет имени В. Ф. Уткина». — Зарегистрировано в Реестре программ для ЭВМ 04.04.2025.

111. Дмитриев, В. Т. Программа моделирования системы передачи информации при восстановлении сигнала по его прореженным отсчётам и отсчётам его трёх первых производных : свидетельство о государственной регистрации программы для ЭВМ № 2025619213 / В. Т. Дмитриев, Ву Хоанг Шон; правообладатель ФГБОУ ВО «Рязанский государственный радиотехнический университет имени В. Ф. Уткина». — Зарегистрировано в Реестре программ для ЭВМ 14.04.2025.

112. Дмитриев, В. Т. Программа моделирования многоэтапного разложения речевых сигналов на основе модификации алгоритма Хургина – Яковлева: свидетельство о государственной регистрации программы для ЭВМ № 2025663352 / В. Т. Дмитриев, Ву Хоанг Шон; правообладатель ФГБОУ ВО «Рязанский государственный радиотехнический университет имени В. Ф. Уткина». — Зарегистрировано в Реестре программ для ЭВМ 27.05.2025.

113. Дмитриев, В. Т. Программа сжатия речевых сигналов на основе энергетических порогов в дискретном вейвлет-преобразовании с оптимизацией методом роя частиц : свидетельство о государственной регистрации программы для ЭВМ № 2026611916 / В. Т. Дмитриев, Ву Хоанг Шон; правообладатель ФГБОУ ВО «Рязанский государственный радиотехнический университет имени В. Ф. Уткина». — Зарегистрировано в Реестре программ для ЭВМ 22.01.2026.

114. Ву Хоанг Шон Актуальность разработки низкоскоростных первичных кодеков речевых сигналов // Новые информационные технологии в научных

исследованиях и образовании: материалы XXVIII Всероссийской НТК студентов, молодых учёных и специалистов «НИТ-2024». — Рязань: РГРТУ, 2023. — С. 138–140.

115. Дмитриев, В. Т. Модификация алгоритма Хургина – Яковлева при трёхканальной реализации / В. Т. Дмитриев, Ву Хоанг Шон // Современные технологии в науке и образовании : сборник трудов VII Международного НТФ «СТНО-2024» : в 10 т. — Рязань : РГРТУ, 2024. — Т. 1. — С. 139–143.

116. Дмитриев, В. Т. Сравнение эффективности первичных кодеков речевых сигналов на основе модификации алгоритма Хургина – Яковлева / В. Т. Дмитриев, Ву Хоанг Шон // Электронные средства и системы управления: материалы XX Международной НПК «ЭСиСУ-2024». — Томск: В-Спектр, ТУСУР, 2024. — Ч. 1. — С. 33–35.

117. Дмитриев, В. Т. Восстановление речевых сигналов на основе модифицированного алгоритма Хургина – Яковлева при четырёхканальной реализации / В. Т. Дмитриев, Ву Хоанг Шон // Актуальные проблемы и перспективы развития радиотехнических и инфокоммуникационных систем: сборник научных статей по материалам VIII Международной НПК «РАДИОИНФОКОМ-2024». — М.: МИРЭА, 2024. — С. 159–164.

118. Ву Хоанг Шон Выигрыш качества воспроизведения речевых сигналов с алгоритмом модифицированного Хургина – Яковлева в трёх- и четырёхканальных реализациях // Новые информационные технологии в научных исследованиях и образовании: материалы XXIX Всероссийской НТК студентов, молодых учёных и специалистов «НИТ-2024». — Рязань: РГРТУ, 2024. — С. 132–134.

119. Дмитриев, В. Т. Разработка и исследование помехоустойчивой системы передачи речевой информации на основе алгоритма Хургина – Яковлева / В. Т. Дмитриев, Ву Хоанг Шон // Актуальные проблемы и перспективы развития радиотехнических и инфокоммуникационных систем: сборник научных статей по материалам VIII Международной НПК «АПСНИП-2024». — Рязань: РГРТУ, 2024. — С. 51–58.

120. Ву Хоанг Шон Анализ вычислительных затрат при обработке речевых сигналов на основе алгоритма Хургина – Яковлева // Промышленная революция 4.0:

взгляд молодёжи: тезисы докладов 6-й Межрегиональной научной сессии молодых исследователей. — Тула: ТулГУ, 2024. — С. 32–33.

121. Дмитриев, В. Т. Определение выигрыша помехоустойчивости многоканальной системы передачи на основе модификации алгоритма Хургина – Яковлева / В. Т. Дмитриев, Ву Хоанг Шон // Информационные технологии. Радиоэлектроника. Телекоммуникации: материалы XI Международной заочной НТК «ITRT-2024». — Тольятти: ПВГУС, 2025. — С. 445–451.

122. Дмитриев, В. Т. Система обработки и передачи информации на основе модификации алгоритма Хургина – Яковлева при повторном разложении / В. Т. Дмитриев, Ву Хоанг Шон // Современные технологии в науке и образовании : сборник трудов VIII Международного НТФ «СТНО-2025» : в 10 т. — Рязань: РГРТУ, 2025. — Т. 1. — С. 171–176.

123. Дмитриев, В. Т. Сравнение помехоустойчивости радиосистем на основе модификации алгоритма Хургина – Яковлева и теоремы В. А. Котельникова / В. Т. Дмитриев, Ву Хоанг Шон // Актуальные проблемы радиоэлектроники и телекоммуникаций: материалы Всероссийской НТК. — Самара: ООО «АРТЕЛЬ», 2025. — С. 184–186.

124. Дмитриев, В. Т. Сжатие речевых сигналов на основе вейвлет-пакетного разложения / В. Т. Дмитриев, Ву Хоанг Шон // Радиотехника, электроника и связь: труды VIII Международной НТК «РЭиС-2025». — Омск, 2025. — С. 184–186.

125. Дмитриев, В. Т. Оценка эффективности кодеков речи на основе вейвлет-пакетного разложения / В. Т. Дмитриев, Ву Хоанг Шон // Электронные средства и системы управления: материалы XXI Международной НПК «ЭСиСУ-2025». — Томск: В-Спектр, ТУСУР, 2025. — Ч. 1. — С. 33–35.

126. Ву Хоанг Шон Алгоритмы пороговой обработки в вейвлет-разложении для сжатия речевых сигналов // Новые информационные технологии в научных исследованиях и образовании: материалы XXX Юбилейной Всероссийской НТК студентов, молодых учёных и специалистов «НИТ-2025». — Рязань: РГРТУ, 2025. — С. 125–126.

127. Дмитриев, В. Т. Метод сжатия речевых сигналов на основе дискретного

вейвлет-преобразования с энергетическим пороговым отсечением / В. Т. Дмитриев, Ву Хоанг Шон // Актуальные проблемы и перспективы развития радиотехнических и инфокоммуникационных систем : сборник научных статей по материалам VIII Международной НПК «АПСНИП-2025». — Рязань: РГРТУ, 2025. — С. 51–58.

128. Дмитриев, В. Т. Метод первичного кодирования речи на основе модифицированного алгоритма Хургина – Яковлева и дискретного вейвлет-преобразования / В. Т. Дмитриев, Ву Хоанг Шон // Современные технологии в науке и образовании : сборник трудов IX Международного НТФ «СТНО-2026» : в 10 т. — Рязань: РГРТУ, 2025. — Т. 1. — С. 171–176.

129. Дмитриев, В. Т. Экспериментальная оценка эффективности алгоритмов первичного кодирования речевых сигналов на основе представления Хургина – Яковлева / В. Т. Дмитриев, Ву Хоанг Шон // Радиолокация, навигация, связь : сборник трудов XXXII Международной НТК : в 6 т. — Воронеж: Издательский дом ВГУ, 2026. — Т. 1. — С. 171–176.

130. Зюко, А. Г. Помехоустойчивость и эффективность систем передачи информации / А. Г. Зюко, А. И. Фалько, И. П. Панфилов [и др.]. — М.: Радио и связь, 1985. — 272 с.

131. Скляр, Б. Цифровая связь. Теоретические основы и практическое применение / Пер. с англ. — 2-е изд. — М.: Вильямс, 2007. — 1104 с. — ISBN 978-5-8459-0498-6.

132. Попов, О. Б. Цифровая обработка сигналов в трактах звукового вещания / О. Б. Попов, С. Г. Рихтер. — М.: Горячая линия–Телеком, 2007. — 240 с. — ISBN 978-5-9912-0018-2.

133. Шувалов, В. П. Телекоммуникационные системы и сети : учебник / В. П. Шувалов, В. В. Захарченко, В. О. Шварцман. — М.: Горячая линия–Телеком, 2010. — Т. 3. — 592 с. — ISBN 978-5-9912-0152-3.

134. Фарков, Ю. А. Биортогональные вейвлеты и фреймы на конечных полях // Математические заметки. — 2009. — Т. 86, № 5. — С. 741–756. — DOI: 10.4213/mzm8574.

СЛОВАРЬ СОКРАЩЕНИЙ

АБГШ	—	аддитивный белый гауссовский шум
АДИКМ	—	адаптивная дифференциальная импульсно-кодовая модуляция
АЦП	—	аналого-цифровой преобразователь
АЧХ	—	амплитудно-частотная характеристика
БПФ	—	быстрое преобразование Фурье
ДВП	—	дискретное вейвлет-преобразование
ДЕКОД	—	декодер (блок декодирования в структурной схеме)
ИКМ	—	импульсно-кодовая модуляция
ИПХ	—	импульсная переходная характеристика
КВ	—	квантователь (блок квантования в структурной схеме)
КОД	—	кодер (блок кодирования в структурной схеме)
КС	—	канал связи
НСКО	—	нормированная среднеквадратическая ошибка
ОДВП	—	обратное дискретное вейвлет-преобразование
ОСШ	—	отношение сигнал/шум
РС	—	речевой сигнал
СКО	—	среднеквадратическое отклонение
СПМ	—	спектральная плотность мощности
ФНЧ	—	фильтр нижних частот
ФЧХ	—	фазочастотная характеристика
ЦАП	—	цифро-аналоговый преобразователь
ACELP	—	Algebraic Code-Excited Linear Prediction (алгебраическое кодово-возбуждаемое линейное предсказание)
LBRAMR	—	Low-Bitrate Adaptive Multi-Rate (низкоскоростной адаптивный многоскоростной кодек)
MAC	—	Multiply-Accumulate (операция умножения-накопления)
MELP	—	Mixed Excitation Linear Prediction (линейное предсказание со смешанным возбуждением)
MIPS	—	Million Instructions Per Second (миллионов инструкций в секунду)

MOS — Mean Opinion Score (средняя экспертная оценка качества речи по шкале МСЭ-Т)

MSE — Mean Squared Error (среднеквадратическая погрешность восстановления)

PSO — Particle Swarm Optimization (оптимизация роем частиц)

SegSNR — Segmental Signal-to-Noise Ratio (сегментное отношение сигнал/шум)

SNR — Signal-to-Noise Ratio (отношение сигнал/шум)

SUREShrink — Stein's Unbiased Risk Estimate Shrinkage

ViSQOL — Virtual Speech Quality Objective Listener (объективная перцептивная метрика качества речи)

VisuShrink — Visual Shrinkage (универсальный пороговый метод Донохо–Джонстона)

ПРИЛОЖЕНИЕ А. ВЫЧИСЛИТЕЛЬНАЯ СЛОЖНОСТЬ СИСТЕМЫ НА ОСНОВЕ ПРЕДСТАВЛЕНИЯ ХУРГИНА –- ЯКОВЛЕВА

Расчёт вычислительной сложности системы на основе представления Хургина – Яковлева выполняется путём раздельного подсчёта операций вещественного умножения и вещественного сложения. Вычислительная единица — вещественная арифметическая операция (умножение × или сложение +). Бабочка БПФ Кули–Тьюки радикс-2 выполняет 1 комплексное умножение (4 вещественных умножения и 2 вещественных сложений) и 2 комплексных сложения (4 вещественных сложений), итого 4× + 6+ на одну бабочку. Для БПФ(N) длиной N имеется $(N/2)\log_2 N$ бабочек, откуда $2N\log_2 N$ умножений и $3N\log_2 N$ сложений. ОБПФ той же длины имеет идентичную схему бабочек и ту же стоимость. Для одного речевого канала при частоте дискретизации $f_{\text{Д}} = 8000$ Гц и длине кадра $M = 256$ частота следования кадров $f_{\text{кадр}} = f_{\text{Д}}/M = 31{,}25$ кадр/с, а требуемая производительность в MIPS: $\text{П} = C_{\Sigma} f_{\text{кадр}}/10^6$, где C_{Σ} — суммарное число вещественных арифметических операций за один кадр.

Рассмотрим блок анализа с K каналами и длиной кадра M при однократном разложении ($P = 1$). Вход — M вещественных отсчётов за один кадр; выход — K каналов по M/K децимированных отсчётов.

Шаг 1. Прямое БПФ(M) — однократно, общий для K каналов. Формируется общий спектр $X[m]$ для всех K каналов; вычисляется один раз независимо от K, что и обеспечивает основной вычислительный эффект системы на основе представления Хургина – Яковлева по сравнению с K независимыми БПФ:

$$N_{\times,1} = 2M\log_2 M,\ N_{+,1} = 3M\log_2 M.$$

Шаг 2a. Канал $k = 0$: оператор $(j2\pi f)^0 = 1$. Нулевой канал передаёт исходный сигнал без дифференцирования; умножение на единицу не требует арифметических операций: $N_{\times,2a} = 0,\ N_{+,2a} = 0.$

Шаг 2b. Каналы $k = 1,\ldots,K-1$: операторы дифференцирования. Для каждого из $(K-1)$ каналов выполняется M поэлементных умножений $X[m]$ на

предвычисленные коэффициенты $W_k[m]$, хранящиеся в таблице. Одно комплексное умножение разворачивается в 4 вещественных умножения и 2 вещественных сложения: $N_{\times,2б} = 4(K-1)M$, $N_{+,2б} = 2(K-1)M$.

Шаг 3. K × ОБПФ(M) — до децимации. Каждый из K каналов требует обратного преобразования длиной M для восстановления временных отсчётов производных в полном разрешении до децимации. Это наиболее дорогостоящий шаг: $N_{\times,3} = K.2M\log_2 M$, $N_{+,3} = K.3M\log_2 M$.

Шаг 4. Децимация ↓K. Из M отсчётов каждого канала выбирается каждый K-й отсчёт — адресная операция без арифметики: $N_{\times,4} = 0$, $N_{+,4} = 0$.

Суммируя вклады всех четырёх шагов и группируя члены при $\log_2 M$:

$$C_{A,\times} = 2\left(K+1\right)M\log_2 M + 4(K-1)M\,, \tag{A.1}$$

$$C_{A,+} = 3\left(K+1\right)M\log_2 M + 2(K-1)M\,, \tag{A.2}$$

$$C_A = C_{A,\times} + C_{A,+} = 5\left(K+1\right)M\log_2 M + 6(K-1)M. \tag{A.3}$$

Блок синтеза при $P=1$ принимает K каналов по M/K децимированных отсчётов. Для каждого канала k синтезирующий фильтр Φ_k реализуется посредством ОБПФ(M/K): $\frac{M/K}{2}\log_2\left(M/K\right)$ бабочек на один канал, суммарно для K каналов:

$$N_{\times,\text{фильтр}} = 4K\frac{M/K}{2}\log_2\frac{M}{K} = 2M\log_2\frac{M}{K},$$

$$N_{+,\text{фильтр}} = 6K\frac{M/K}{2}\log_2\frac{M}{K} = 3M\log_2\frac{M}{K}.$$

Суммирование K восстановленных последовательностей добавляет $\left(K-1\right)$ сложений на каждый из M отсчётов:

$$C_{S,\times} = 2M\log_2\frac{M}{K}, \tag{A.4}$$

$$C_{S,+} = 3M\log_2\frac{M}{K} + (K-1)M, \tag{A.5}$$

$$C_S = C_{S,\times} + C_{S,+} = 5M\log_2\frac{M}{K} + (K-1)M. \tag{A.6}$$

Обобщение на P этапов. На этапе $p \in \{1,\ldots,P\}$ входной сигнал разбит на K^{p-1} независимых групп, каждая из которых содержит $M_P = M/K^{p-1}$ отсчётов. Основной инвариант: суммарное число отсчётов на каждом этапе равно M, поскольку $K^{p-1}.\dfrac{M}{K^{p-1}} = M$. Вследствие этого инварианта структура вычислений этапа p точно воспроизводит структуру $P=1$ с заменой M на M/K^{p-1}, и масштабирующий множитель K^{p-1} сокращается. Стоимость этапа p анализа:

$$C_A^{(p)} = 5\left(K+1\right)M\log_2\frac{M}{K^{p-1}} + 6(K-1)M. \tag{A.7}$$

Подставляя $\log_2\left(M/K^{p-1}\right) = \log_2 M - (p-1)\log_2 K$ и суммируя по $p=1,\ldots,P$:

$$\sum_{p=1}^{P}\log_2\frac{M}{K^{p-1}} = P\log_2 M - \log_2 K\sum_{p=1}^{P}\left(p-1\right) = P\log_2 M - \log_2 K\frac{P(P-1)}{2},$$

где использована формула суммы $\sum_{p=1}^{P}(p-1) = P(P-1)/2$. Итого:

$$C_{A,\times} = 2\left(K+1\right)M\left[P\log_2 M - \log_2 K\frac{P(P-1)}{2}\right] + 4(K-1)MP, \tag{A.8}$$

$$C_{A,+} = 3\left(K+1\right)M\left[P\log_2 M - \log_2 K\frac{P(P-1)}{2}\right] + 2(K-1)MP, \tag{A.9}$$

$$C_A(K,M,P) = 5\left(K+1\right)M\left[P\log_2 M - \log_2 K\frac{P(P-1)}{2}\right] + 6(K-1)MP. \tag{A.10}$$

Аналогично для синтеза — на этапе p имеется K^{p-1} групп, в каждой K подканалов с M/K^{p} отсчётами, применяется ОБПФ(M/K^{p}). Суммируя с $\sum_{p=1}^{P} p = P(P+1)/2$:

$$C_S(K,M,P) = 5M\left[P\log_2 M - \log_2 K\frac{P(P+1)}{2}\right] + (K-1)MP. \tag{A.11}$$

Численная верификация для $K=2$, $P=1$, $M=256$. Пошаговое распределение операций приведено в таблице А.1.

Таблица А.1. Пошаговое распределение операций при $K=2$, $P=1$, $M=256$

№	Шаг	Обоснование	×	+	Итого
	Блок анализа				
1	БПФ(256): 1 раз	1024 бабочки	4096	6144	10240
2a	Канал $k=0$: оператор $(j2\pi f)^0=1$	арифметики нет	0	0	0
2b	Канал $k=1$: оператор дифференцирования	256 комплексных умножений	1024	512	1536
3	2 × ОБПФ(256)	2048 бабочек	8192	12288	20480
4	Децимация ↓2	Выборка M/K из M: операция памяти			
Σ	C_A итого		13312	18944	32256
	Блок синтеза				
1	Интерполяция ↑2	Вставка $K-1$ нулей между отсчётами: операция памяти	0	0	0
2	2 × ОБПФ(128) — синтезирующий фильтр	896 бабочек	3584	5376	8960
3	Суммирование 2 канала	$M(K-1)=256$ сложений	0	256	256
Σ	C_S итого		3584	5632	9216

Сравнение с системой В.А. Котельникова. Одноканальная система В.А. Котельникова ($K=1$) реализует полный цикл БПФ(M) + ОБПФ(M): $C_{КОТ}=10M\log_2 M=10.256.8=20480$ оп. при $M=256$. Требуемая производительность: $\Pi_{КОТ}=\dfrac{20480.31,25}{10^6}=0,64$ MIPS. В таблице А.2 для каждой конфигурации системы на основе представления Хургина – Яковлева приведены суммарная производительность $\Pi_{ХЯ}$ и разность $\Delta\Pi=\Pi_{ХЯ}-\Pi_{КОТ}$, показывающая абсолютный прирост нагрузки по отношению к одноканальному ориентиру.

Таблица А.2. Вычислительная сложность конфигураций системы на основе представления Хургина – Яковлева и сравнение с системой В.А. Котельникова по требуемой производительности

K, P	K^P	C_Σ, оп.	П, MIPS	ΔП, MIPS
K=1, P=1	1	20480	0,640	0
K=2, P=1	2	41472	1,296	+0,656
K=2, P=2	4	77824	2,432	+1,792
K=2, P=3	8	109056	3,408	+2,768
K=3, P=1	3	52755	1,649	+1,008
K=3, P=2	9	95365	2,980	+2,340
K=3, P=3	27	127834	3,995	+3,355
K=4, P=1	4	64256	2,008	+1,368
K=4, P=2	16	113152	3,536	+2,896
K=4, P=3	64	146688	4,584	+3,944

Из таблицы А.2 следует, что разность ΔП положительна для всех конфигураций: система на основе представления Хургина – Яковлева всегда требует большей производительности, чем одноканальная система В.А. Котельникова. Это закономерно, поскольку система на основе представления Хургина – Яковлева одновременно формирует K^P аналитических каналов. Содержательной мерой эффективности служит удельная стоимость одного канала Π / K^P: при K = 2, P = 1 она составляет 1,296/2=0,6481 MIPS/канал — на 1,25 % выше стоимости одного канала В.А. Котельникова (0,640 MIPS). Уже при $K = 2$, $P = 2$ удельная стоимость снижается до 2,432/4=0,608 MIPS/канал, то есть оказывается ниже уровня В.А. Котельникова; при $K = 3$, $P = 1$ — 1,649/3=0,550 MIPS/канал; при $K = 4$, $P = 1$ — 2,008/4=0,502 MIPS/канал. Таким образом, удельная вычислительная стоимость канала в системе на основе представления Хургина – Яковлева монотонно убывает с ростом K и P, что подтверждает фундаментальное вычислительное преимущество многоканального представления Хургина – Яковлева при масштабировании числа аналитических каналов.

ПРИЛОЖЕНИЕ Б. КОПИЯ АКТОВ ВНЕДРЕНИЯ

«Утверждаю»
Проректор по образовательной деятельности
ФГБОУ ВО «РГРТУ»
_____ Е.А. Соколова
«21» февраля 2026 г.

АКТ
о внедрении результатов диссертационной работы
Ву Хоанг Шон
в учебный процесс ФГБОУ ВО «Рязанский государственный радиотехнический университет имени В.Ф. Уткина»
(ФГБОУ ВО «РГРТУ»)

Настоящий акт составлен о том, что материалы диссертации на соискание учёной степени кандидата технических наук Ву Хоанг Шон используются в учебном процессе подготовки бакалавров по направлению 11.03.02 «Инфокоммуникационные технологии и системы связи» и магистрантов по направлению 11.04.02 «Инфокоммуникационные технологии и системы связи».

Внедрены следующие материалы диссертационной работы Ву Хоанг Шон:

1. Результаты исследования обобщённой дискретизации на основе представления Хургина – Яковлева и его сочетания с дискретным вейвлет-преобразованием включены в лекционный курс по дисциплине «Методы первичного кодирования в телекоммуникациях» для магистрантов по направлению 11.04.02.

2. Разработанные алгоритмы кодирования речевых сигналов на основе дискретного вейвлет-преобразования с адаптивной настройкой энергетических порогов методом роя частиц используются в лабораторных работах по дисциплине «Кодеки сигналов в МТКС» для бакалавров по направлению 11.03.02, что способствует формированию у студентов практических навыков в области современных методов обработки и сжатия речи в условиях ограниченных ресурсов канала связи.

Декан факультета радиотехники и телекоммуникаций _____ Егоров А.В.

Председатель методической комиссии факультета радиотехники и телекоммуникаций _____ Корнеев В.А.

Заведующий кафедрой РУС _____ Дмитриев В.Т.

УТВЕРЖДАЮ
Директор филиала
ПАО «МТС» в г. Рязани

_____Ляк А.Л.

«____» _____ 2025 г.

АКТ

о внедрении результатов кандидатской диссертационной работы

Ву Хоанг Шон

в ПАО «Мобильные ТелеСистемы» филиал г. Рязань

Комиссия в составе:

председателя:

директор филиала ПАО «МТС» в г. Рязань Ляк А. Л.;

члены комиссии:

технический директор филиала ПАО «МТС» в г. Рязань Смирнягин В. В.;

начальник отдела развития сети ПАО «МТС» в г. Рязань Мороз В. Ю.

составила настоящий акт о том, что в 2025 г. следующие результаты диссертационной работы Ву Хоанг Шон, представленной на соискание ученой степени кандидата технических наук были внедрены в ПАО «Мобильные ТелеСистемы» филиал г. Рязань:

1. Алгоритмы первичного кодирования речи на основе трёх- и четырёхканальных модификаций алгоритма Хургина – Яковлева, обеспечивающие повышение помехоустойчивости системы в условиях ошибок канала связи.
2. Обобщённый подход к параллельной обработке речевых сигналов на основе модифицированного алгоритма Хургина – Яковлева с многоэтапным

разложением, обеспечивающий повышенное качество восстановления и снижение вычислительной сложности.

3. Комбинированный алгоритм первичного кодирования речи, объединяющий преимущества модификации алгоритма Хургина – Яковлева и дискретного вейвлет-преобразования для повышения общего качества восстановления речи.

Предложенные в диссертации Ву Хоанг Шон алгоритмы первичного кодирования речевых сигналов, основанные на трёх- и четырёхканальных модификациях алгоритма Хургина – Яковлева, обеспечивают превосходство над системами, построенными на основе теоремы В. А. Котельникова, в условиях ошибок канала (вероятность битовых ошибок от 0,1 до 5%). Улучшение качества восстановленной речи составляет: от 0,2 и до 1,4 балла по шкале MOS для низкоскоростных кодеков, от 0,2 и до 1,8 балла — для кодеков средней скорости, до 1,6 балла — для высокоскоростных алгоритмов кодирования.

Применение предложенного алгоритма в двухканальной структуре с двух- или трёхэтапным разложением обеспечивает высокое качество восстановления речи.

Гибридный подход, сочетающий модифицированный алгоритм Хургина – Яковлева и вейвлет-преобразование, обеспечивает эффективный компромисс между качеством восстановления речи и степенью сжатия. Эксперименты показывают, что такое объединение устойчиво превосходит отдельное применение каждого метода, особенно в условиях, где требуется одновременно высокое качество речи и минимальный объём передаваемых данных.

Председатель комиссии: _____ Ляк А. Л.

Члены комиссии: _____ Смирнягин В. В.

_____ Мороз В. Ю.

Публичное акционерное общество «Мобильные ТелеСистемы» Филиал в г.Рязань
ул.Праволыбедская, д.31. г.Рязань, Россия, 390000. Тел.(4912) 40-49-91 Факс: (4912) 90-86-83 www.ryazan.ru

CÔNG TY TNHH TRUYỀN HÌNH FPT

GIẤY CHỨNG NHẬN

Về việc ứng dụng kết quả luận án Tiến sĩ của nghiên cứu sinh Vũ Hoàng Sơn trường "Đại học Tổng hợp Kỹ thuật Vô tuyến Điện tử Ryazan mang tên V.F Utkin" – Liên Bang Nga

Văn bản này xác nhận việc ứng dụng thành công kết quả luận án Tiến sĩ của nghiên cứu sinh Vũ Hoàng Sơn vào xử lý và mã hóa tín hiệu giọng nói số trong môi trường truyền thông thực tế. Phương pháp đề xuất kết hợp thuật toán Xurghin–Yakovlev đã sửa đổi với biến đổi wavelet rời rạc (DWT), trong đó ngưỡng xử lý các hệ số wavelet được tự động tối ưu bằng thuật toán bầy đàn PSO. Giải pháp này đáp ứng đồng thời ba yêu cầu then chốt của hệ thống viễn thông hiện đại — chất lượng thoại cao, khả năng chống lỗi và hiệu quả nén — thông qua phân tích đa độ phân giải và lựa chọn ngưỡng tối ưu theo các tiêu chí SegSNR và ViSQOL, đảm bảo hiệu suất ổn định ngay cả trong điều kiện băng thông hạn chế hoặc lỗi kênh cao (BER 0,1–5%).

Kết quả thực nghiệm cho thấy: khi triển khai phương pháp đạt SNR ≥ 14 дБ, đủ để đảm bảo thoại rõ ràng trong các hệ thống mã hóa tốc độ thấp. Đồng thời, cải thiện chất lượng theo thang đo MOS từ 0,2 đến 1,8 điểm so với các codec chuẩn, tùy theo tốc độ bit. Hiện nay, các thành phần cốt lõi của giải pháp đã được tích hợp thử nghiệm trong các mô-đun xử lý thoại tại **CÔNG TY TNHH TRUYỀN HÌNH FPT**, phục vụ cho việc nâng cao hiệu suất hệ thống truyền giọng nói số trên nền tảng mạng hiện có.

Vì vậy, luận án không chỉ có giá trị học thuật mà còn mang lại hiệu quả ứng dụng rõ rệt, đáp ứng yêu cầu thực tiễn của ngành viễn thông tại Việt Nam trong bối cảnh chuyển đổi số và phát triển hạ tầng thông minh.

Hà Nội, ngày 9 tháng 12 năm 2025

ĐẠI DIỆN CÔNG TY

CÔNG TY TNHH TRUYỀN HÌNH FPT
FPT Telecom

Lương Duy Phương

GIÁM ĐỐC CÔNG NGHỆ

Перевод с вьетнамского языка на русский язык

Общество с ограниченной ответственностью
«FPT Telecom»

АКТ
о внедрении результатов диссертационной работы
Ву Хоанг Шон ФГБОУ ВО «Рязанский государственный радиотехнический университет имени В. Ф. Уткина», Российская Федерация

Настоящий документ подтверждает успешное внедрение результатов диссертации соискателя Ву Хоанг Шон в области цифровой обработки и кодирования речевых сигналов в условиях реальных телекоммуникационных систем. Предложенный метод основан на комбинации модифицированного алгоритма Хургина – Яковлева и дискретного вейвлет-преобразования (DWT), при этом пороговая обработка вейвлет-коэффициентов автоматически оптимизируется с использованием алгоритма роя частиц (PSO). Данное решение одновременно удовлетворяет трём ключевым требованиям современных телекоммуникационных систем — высокому качеству речи, устойчивости к ошибкам канала и эффективности сжатия — за счёт многоуровневого анализа сигнала и выбора оптимальных порогов по критериям SegSNR и ViSQOL, обеспечивая стабильную производительность даже при ограниченной полосе пропускания или высоком уровне битовых ошибок (BER 0,1–5 %).

Экспериментальные результаты показали, что при реализации данного метода достигается отношение сигнал/шум (SNR) не менее 14 дБ, что достаточно для обеспечения разборчивой речи в системах низкоскоростного кодирования. Одновременно достигается улучшение качества речи по шкале MOS на 0,2–1,8 балла по сравнению со стандартными кодеками (в зависимости от битрейта). В настоящее время ключевые компоненты разработанного решения проходят опытную интеграцию в модули речевой обработки компании **ООО «FPT Telecom»** для повышения эффективности существующих систем передачи речи.

Таким образом, диссертация представляет не только научную, но и значимую практическую ценность, отвечая актуальным потребностям отрасли связи во Вьетнаме в условиях цифровой трансформации и развития интеллектуальной инфраструктуры.

Ханой, «09» декабря 2025 года
Директор По Технологиям
/Печать: ООО «FPT Telecom»/
Луонг Дуй Фыонг
(Подписано и поставлена печать)

Перевод данного текста выполнен переводчиком Васюковой Екатериной Юрьевной

Российская Федерация

Город Москва

Десятого декабря две тысячи двадцать пятого года

Я, Карнаухова Айгуль Энгельсовна, нотариус города Москвы, свидетельствую подлинность подписи переводчика Васюковой Екатерины Юрьевны.

Подпись сделана в моем присутствии.

Личность подписавшего документ установлена.

Зарегистрировано в реестре: № 77/236-н/77-2025- 25 - 311

Уплачено за совершение нотариального действия: 400 руб. 00 коп.

А.Э. Карнаухова

Пронумеровано, прошнуровано
и скреплено печатью 2 листа (ов)

Нотариус

ПРИЛОЖЕНИЕ В. КОПИЯ СВИДЕТЕЛЬСТВ О ГОСУДАРСТВЕННОЙ РЕГИСТРАЦИИ ПРОГРАММЫ ДЛЯ ЭВМ

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2024619863

«Программа моделирования модификации алгоритма Хургина - Яковлева при трёхканальной реализации»

Правообладатель: ***федеральное государственное бюджетное образовательное учреждение высшего образования "Рязанский государственный радиотехнический университет имени В.Ф. Уткина" (RU)***

Авторы: ***Дмитриев Владимир Тимурович (RU), Ву Хоанг Шон (VN)***



Заявка № **2024617055**

Дата поступления **04 апреля 2024 г.**

Дата государственной регистрации в Реестре программ для ЭВМ ***27 апреля 2024 г.***

Руководитель Федеральной службы по интеллектуальной собственности

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2025618510

«Программа для исследования синтезирующих фильтров Хургина-Яковлева»

Правообладатель: ***Федеральное государственное бюджетное образовательное учреждение высшего образования «Рязанский государственный радиотехнический университет имени В.Ф. Уткина» (RU)***

Авторы: ***Дмитриев Владимир Тимурович (RU), Ву Хоанг Шон (VN)***

Заявка № **2025616339**

Дата поступления **21 марта 2025 г.**

Дата государственной регистрации в Реестре программ для ЭВМ ***04 апреля 2025 г.***

Руководитель Федеральной службы по интеллектуальной собственности

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2025619213

«Программа моделирования системы передачи информации при восстановлении сигнала по его прореженным отсчетам и отсчетам его трех первых производных»

Правообладатель: ***федеральное государственное бюджетное образовательное учреждение высшего образования «Рязанский государственный радиотехнический университет имени В.Ф. Уткина» (RU)***

Авторы: ***Дмитриев Владимир Тимурович (RU), Ву Хоанг Шон (VN)***

Заявка № **2025616352**

Дата поступления **21 марта 2025 г.**

Дата государственной регистрации в Реестре программ для ЭВМ ***14 апреля 2025 г.***

Руководитель Федеральной службы по интеллектуальной собственности

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2025663352

Программа моделирования многоэтапного разложения речевых сигналов на основе модификации алгоритма Хургина-Яковлева

Правообладатель: ***Федеральное государственное образовательное учреждение высшего образования "Рязанский государственный радиотехнический университет имени В.Ф. Уткина" (RU)***

Авторы: ***Дмитриев Владимир Тимурович (RU), Ву Хоанг Шон (VN)***

Заявка № **2025661208**

Дата поступления **06 мая 2025 г.**

Дата государственной регистрации в Реестре программ для ЭВМ ***27 мая 2025 г.***

Руководитель Федеральной службы по интеллектуальной собственности

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ
Сертификат 0692e7c1a6300bf54f240f670bcca2026
Владелец Зубов Юрий Сергеевич
Действителен с 10.07.2024 по 03.10.2025

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ



ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

RU2026611916

ГОСУДАРСТВЕННАЯ РЕГИСТРАЦИЯ ПРОГРАММЫ ДЛЯ ЭВМ

Номер регистрации (свидетельства):
2026611916
Дата регистрации: **22.01.2026**
Номер и дата поступления заявки:
2026611004 22.01.2026
Дата публикации и номер бюллетеня:
22.01.2026 Бюл. № 2

Автор(ы):
Дмитриев Владимир Тимурович (RU),
Ву Хоанг Шон (VN)
Правообладатель(и):
Федеральное государственное бюджетное образовательное учреждение высшего образования «Рязанский государственный радиотехнический университет имени В.Ф. Уткина» (RU)

Название программы для ЭВМ:
Программа сжатия речевых сигналов на основе энергетических порогов в дискретном вейвлет-преобразовании с оптимизацией методом роя частиц

Реферат:
Программа реализует сжатие речевых сигналов на основе энергетических порогов в дискретном вейвлет-преобразовании, где параметры порогов оптимизируются методом роя частиц PSO. Программа использует раздельные энергетические пороги для аппроксимирующих и детализирующих коэффициентов каждого уровня разложения. Программа предназначена для применения в системах кодирования речи с ограниченной пропускной способностью, а также для научных исследований в области разреженного представления сигналов.

Язык программирования: MatLab R2024a

Объем программы для ЭВМ: 28 Кб

ПРИЛОЖЕНИЕ Г. КОПИЯ ДИПЛОМОВ



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
ИМЕНИ В.Ф. УТКИНА

ДИПЛОМ

НАГРАЖДАЕТСЯ

Ву Хоанг Шон

за лучший доклад

на XXIX Всероссийской
научно-технической конференции
студентов, молодых ученых и специалистов

***«Новые информационные технологии
в научных исследованиях»***

И.о. ректора С.А. Банников

Рязань 28 ноября 2024г.

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ «РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
ИМЕНИ В.Ф. УТКИНА»

ДИПЛОМ

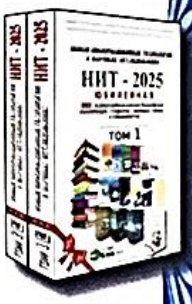
НАГРАЖДАЕТСЯ

Ву Хоанг Шон

за лучший доклад

на **XXX** Юбилейной Всероссийской
научно-технической конференции
студентов, молодых учёных и специалистов

***«Новые информационные технологии
в научных исследованиях»***



И.о. ректора С.А. Банников

Рязань 27 ноября 2025г.




ДИПЛОМ
за лучший доклад
на X Всероссийской
научно-технической конференции
Актуальные проблемы
современной науки и производства
АПСНИП- 2025
награждается
Ву Хоанг Шон
Проректор
С.И. Гусев


ДИПЛОМ
3 степени
за лучший доклад
на VIII Международном
научно-техническом форуме
Современные технологии
в науке и образовании
СТНО - 2025
награждается
Дмитриев
Владимир Тимурович,
Ву Хоанг Шон
Проректор
С.И. Гусев