

На правах рукописи



Фадеев Даниил Романович

**ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ КОДИРОВАНИЯ  
КОЭФФИЦИЕНТОВ ВЕЙВЛЕТНОГО ПРЕОБРАЗОВАНИЯ  
В КОДЕКАХ С КОМПРЕССИЕЙ ЦИФРОВЫХ АУДИОДАНЫХ**

Специальность

05.12.04-Радиотехника, в том числе системы и устройства телевидения

**Автореферат**

диссертации на соискание ученой степени  
кандидата технических наук

Санкт-Петербург – 2017

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича» (СПбГУТ) на кафедре «Радиосвязи и вещания».

**Научный руководитель:**

**Ковалгин Юрий Алексеевич**

заслуженный деятель науки РФ, доктор технических наук, профессор, Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича» (СПбГУТ), профессор кафедры «Радиосвязи и вещания»

**Официальные оппоненты:**

**Кривошейкин Анатолий Валентинович**

доктор технических наук, профессор, «Санкт-Петербургский государственный институт кино и телевидения» (СПбГИКиТ), профессор кафедры «Радиотехники и информационных технологий»

**Чесноков Михаил Александрович**

кандидат технических наук, Акционерное общество «Научно-технический центр «Ядерно-физические исследования» (АО «НТЦ «ЯФИ»), ведущий инженер-электроник

**Ведущая организация:**

ЗАО «Проектно-конструкторское бюро «РИО»  
(ЗАО «ПКБ «РИО»), г. Санкт-Петербург

Защита состоится «21» июня 2017 г. в 16.00 часов на заседании диссертационного совета Д 212.238.03 при Санкт-Петербургском государственном электротехническом университете «ЛЭТИ» им. В. И. Ульянова (Ленина) по адресу: 197376, Санкт-Петербург, ул. Проф. Попова, дом 5., ауд. 2316.

С диссертацией можно ознакомиться в библиотеке университета и на сайте <http://www.eltech.ru/>.

Автореферат разослан «20» апреля 2017 г.

Ученый секретарь  
диссертационного совета Д 212.238.03  
к.т.н., доцент



М. Е. Шевченко

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы диссертации.** В последнее десятилетие весьма отчетливо проявились очертания грядущей информационной эпохи. Новые сетевые сервисы, предоставляемые пользователю, порождают новые запросы и требования к качеству мультимедийного контента. Если в конце 20 века доступ к Интернет осуществлялся со скоростями в 10-ки кБайт в секунду, что само по себе исключало возможность воспроизведения звука высокого качества из сети в режиме реального времени, то в настоящее время большинство провайдеров сотовой связи обеспечивают уверенное 3G/4G-покрытие в крупных городах России.

На этом фоне алгоритмы сжатия аудиоданных и сегодня не теряют своей актуальности. Изначально разработанные для более эффективного использования носителей информации алгоритмы сжатия звука с потерями в настоящее время находят новые сферы применения. Необходимость передачи мультимедийной информации в условиях работы новых протоколов, ориентированных на сети пост-NGN (Next Generation Networks), обуславливает как использование существующих алгоритмов сжатия, так и разработку новых, ориентированных на конкретные приложения. Широкое применение они находят также и в системах цифрового телевидения и радиовещания, в системах мобильной связи, различного рода мобильных устройствах.

Значительный вклад в развитие теории и практики психоакустики и компрессии цифровых аудиоданных внесли работы таких авторов как: E. Zwicker, R. Plomp, J. Blauert, B. Moore, J. Johnston, K. Brandenburg, E. Terhard, R. Kapust, M. Bosi, F. Spanias, И.А. Алдошина, Ю.А. Ковалгин, Я.Г.Д Придаршан, М.В. Зырянов, И.А. Стефанова, М.А. Стефанов, Г.Г. Рогозинский и ряд других.

Дальнейшая оптимизация алгоритмов сжатия и исследования в этой области являются по-прежнему актуальными. При этом основными направлениями исследований в области алгоритмов сжатия звука с потерями являются модификация ранее используемых или разработка новых моделей психоакустического анализа, оптимизация и разработка новых алгоритмов устранения различных видов избыточности, а также улучшение производительности, качества и эффективности работы кодеков.

**Цель и задачи исследования.** Цель диссертационной работы состоит в разработке кодека с вейвлетным доменом, обеспечивающим повышение эффективности кодирования коэффициентов дискретного вейвлетного преобразования.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Выполнить сравнительный анализ существующих психоакустических моделей, применяемых в алгоритмах компрессии цифровых аудиоданных с потерями.
2. Разработать метод адаптации выбранной психоакустической модели к вейвлетному домену.
3. Исследовать распределение энергии шумов квантования по частоте и разработать метод контроля их уровня при кодировании коэффициентов вейвлетного преобразования.
4. Проверить экспериментально результаты выполненных в работе теоретических исследований.

**Соответствие паспорту специальности.** Содержание исследования соответствует следующим пунктам специальности 05.12.04 – Радиотехника, в том числе системы и устройства телевидения:

*Пункт 6.* Исследование и разработка радиотехнических систем и устройств передачи информации, в том числе радиорелейных и телеметрических, с целью повышения их пропускной способности и помехозащищенности;

*Пункт 7.* Разработка методов и устройств передачи, приема, обработки, отображения и хранения информации. Разработка перспективных информационных технологий, в том числе цифровых, а также с использованием нейронных сетей для распознавания изображений в радиотехнических устройствах.

Разработаны и исследованы программные модели психоакустического анализа кодеров с компрессией цифровых аудиоданных, а также кодеков, в которых процедуры кодирования высококачественных аудиосигналов выполняются в базе пакетного дискретного вейвлетного преобразования.

*Пункт 10.* Разработка радиотехнических устройств для использования их в промышленности, биологии, медицине, метрологии и др.

Разработана и выполнена оценка качества и эффективности кодеков, работающих в базе дискретного вейвлетного преобразования для развития и совершенствования устройств кодирования аудиоданных, применяемых в медиаиндустрии.

**Методы исследования.** При решении поставленных задач применялись методы спектрального анализа, теории вейвлетов, психоакустики, компьютерного моделирования, математической статистики.

#### **Научная новизна работы.**

1. Оценена эффективность работы психоакустических моделей (ПАМ) стандартов MPEG (ПАМ\_1, ПАМ\_2 и ПАМ\_ААС) и вейвлетной психоакустической модели (ПАМ\_ВП). Доказано, что ПАМ\_ААС наиболее полно учитывает особенности распределения энергии звукового сигнала по частоте, наиболее экономно расходует биты при кодировании звукового сигнала.

2. Разработан метод, позволяющий оценить распределение энергии шумов квантования в кодеке, работающем в вейвлетном пространстве и представить полученные результаты в виде матрицы шумов квантования, содержащей число строк, равное числу субполос кодирования дерева фильтров и число столбцов равное числу полос психоакустического анализа. Данный метод универсален, он пригоден для оценки уровня шумов квантования для анализирующих/синтезирующих банков фильтров любой структуры и сложности в кодеках с компрессией цифровых аудиоданных.

3. Разработан итерационный пошаговый метод контроля энергии шумов квантования, позволяющий, используя матрицу  $\mathbf{M}$ , элементами которой являются значения энергии шумов квантования, психоакустическую модель ПАМ\_ААС и предложенный в работе алгоритм масштабирования энергии шумов квантования, найти требуемую для прозрачного кодирования длину кодовых слов вейвлетных коэффициентов в каждой из субполос кодирования.

4. Реализована в среде MATLAB модель кодека, использующего вейвлетное преобразование сигнала выборки, психоакустическую модель ПАМ\_ААС и алгоритм контроля энергии искажений квантования. Методом субъективно-статистических экспертиз подтверждена эффективность ее работы на реальных звуковых сигналах.

**Достоверность полученных результатов и выводов** подтверждается компьютерным моделированием, опорой на результаты научных трудов ведущих отечественных и мировых ученых в данной области, корректным применением математического аппарата, использованием современных методов статистической обработки полученных результатов, обсуждениями на научных конференциях, результатами субъективно-статистических экспертиз, публикациями результатов работы в рецензируемых научных журналах.

#### **Научные положения, выносимые на защиту.**

1. Характеристики моделей психоакустического анализа, полученные в диссертации, позволяют обосновать возможности применения каждой из них в кодеках с компрессией цифровых аудиоданных в зависимости от предъявляемых к ним требований.

2. Метод получения матрицы шумов квантования вейвлетных коэффициентов позволяет найти реальное распределение их энергии по субполосам психоакустического анализа/кодирования при любой структуре дерева анализирующих/синтезирующих фильтров.

3. Метод контроля энергии шумов квантования позволяет найти минимально возможную длину кодовых слов коэффициентов дискретного вейвлетного преобразования, при которой искажения квантования остаются еще незаметными на слух.

4. Кодек с вейвлетным доменом, содержащий ПАМ\_ААС, матрицу шумов квантования и итерационный алгоритм контроля их уровня обеспечивает прозрачное кодирование при средней длине кодовых слов вейвлетных коэффициентов равной 3,35...4,7 бит при использовании вейвлетного базиса Добеши 10 и 3,71...5,22 бит при использовании вейвлетного базиса Добеши 4.

#### **Практическая значимость.**

1. Разработан аппаратно-программный комплекс на базе MATLAB, позволяющий подключать в виде модулей различные психоакустические модели, исследовать характер распределения бит по субполосам психоакустического анализа, оценивать пороги маскировки, требуемое для прозрачного кодирования аудиоданных число бит, а также автоматизировать процесс обработки полученных при этих исследованиях результатов.

2. Предложенный метод оценки энергии шумов квантования по субполосам кодирования универсален, он пригоден для оценки уровня шумов квантования в субполосах кодирования для анализирующих/синтезирующих банков фильтров любой структуры и сложности.

3. Разработанная программная модель кодека на основе вейвлетного преобразования, психоакустической модели ПАМ\_ААС, а также метод получения матрицы шумов квантования и алгоритм контроля энергии искажений квантования, позволяет оценить его эффективность и определить возможности дальнейшего использования на практике в современных мультимедийных системах инфокоммуникаций.

#### **Реализация и внедрение результатов исследований.**

Результаты диссертационной работы использованы:

– в научно-исследовательском институте телевидения (АО «НИИ телевидения») в 2016-2017 годах при разработке Эскизного и Технического проектов ОКР «Садко» и позволили оптимизировать структуру информационных потоков на оснащаемом локальной телевизионной сетью объектов МО РФ (акт внедрения от «21» марта 2017 г.;

– в Санкт-Петербургском государственном университете телекоммуникаций им. проф. М.А. Бонч-Бруевича при выполнении научно-исследовательских и выпускных работ магистров, обучающихся по направлениям (акт внедрения в учебный процесс от «13» марта 2017 г.):

-11.04.02 – «Инфокоммуникационные технологии и системы связи» профиль «Системы и сети радиосвязи, радиовещания и радиодоступа» при изучении дисциплины «Стандарты, системы и средства цифрового радиовещания» и

-11.04.01 – Радиотехника, профиль «Аудиовидеосистемы и медиакоммуникации» при изучении дисциплины «Формирование, хранение, обработка и передача аудиоконтента в медиакоммуникациях», а также:

Подготовлено учебное пособие для проведения практических занятий, курсового и дипломного проектирования/ Психоакустический анализ в алгоритмах компрессии цифровых аудиоданных. Ковалгин Ю.А., Фадеев Д.Р.; СПбГУТ. –СПб., 2016.

#### **Апробация работы.**

Материалы диссертационной работы были обсуждены на конференциях различного уровня: международной научно-технической конференции, посвященной 125-летию со дня рождения В.И. Волынкина, «Актуальные проблемы радио- и ки-

нотехнологий», СПб.: СПбГИКиТ, 26-28 октября 2016 г.; V Международной научно-технической и научно-методической конференции «Актуальные проблемы инфотелекоммуникаций в науке и образовании», СПб.: СПбГУТ им. проф. М.А. Бонч-Бруевича, 10-11 марта 2016 г.; VI Международной научно-технической и научно-методической конференции «Актуальные проблемы инфотелекоммуникаций в науке и образовании», СПб.: СПбГУТ им. проф. М.А. Бонч-Бруевича, 3-4 марта 2015 г.; III Международной научно-технической и научно-методической конференции «Актуальные проблемы инфотелекоммуникаций в науке и образовании», СПб.: СПбГУТ им. проф. М.А. Бонч-Бруевича, 25-26 февраля 2014 г.; II Международной научно-технической и научно-методической конференции «Актуальные проблемы инфотелекоммуникаций в науке и образовании», СПб.: СПбГУТ им. проф. М.А. Бонч-Бруевича, 26-27 февраля 2013 г.; XVIII международной научно-технической конференции «Приоритетные направления развития науки и технологий», (два доклада), Тула 2015 г.

**Публикации.** Основные теоретические и практические результаты диссертации опубликованы в 9 печатных работах, среди которых 4 научные статьи в рецензируемых изданиях, входящих в перечень рекомендованных ВАК РФ, 5- статей, опубликованных в других изданиях и материалах конференций.

**Личное участие автора в получении научных результатов.** В исследованиях, результаты которых приведены в диссертационной работе, автору принадлежит определяющая роль. Диссертант лично разрабатывал компьютерные программные модели, исследовал приведенные в работе алгоритмы, выполнял анализ и обобщение полученных данных. Часть опубликованных работ по проведенным исследованиям написана в соавторстве с научным руководителем, Заслуженным деятелем науки РФ, д.т.н. профессором Ю.А. Ковалгиным.

**Структура и объем диссертации.** Диссертация состоит из введения, четырех глав с выводами по каждой из них, заключения, выводов, списка литературы и двух приложений.

Диссертация изложена на 152 страницах текста, содержит 65 рисунков, 9 таблиц, список литературы включает в себя 52 отечественных, 110 зарубежных источников.

## **КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ**

**Введение** содержит обоснование актуальности темы диссертационного исследования, формулировку цели и задач работы, положения, выносимые на защиту, сведения о структуре, публикациях и методах исследования, использованных в работе.

**В первой главе** представлено краткое изложение основных принципов, лежащих в основе алгоритмов перцепционного сжатия звука, дан анализ современного состояния проблемы сжатия звука с потерями, обоснована цель исследования, приведен перечень основных задач, решаемых в диссертации.

**Во второй главе** приводится описание и исследование четырех наиболее распространенных психоакустических моделей (ПАМ): ПАМ\_1 и ПАМ\_2 стандартов MPEG ISO/IEC 11172-3, ISO/IEC 13818-3, ПАМ\_AAC стандартов MPEG-2 ISO/IEC 13818-7 AAC, MPEG-4 ISO/IEC 14496-3, а также психоакустической модели, работающей в вейвлетном пространстве (ПАМ\_ВП).

Для исследования ПАМ был разработан в среде MATLAB аппаратно-программный комплекс. В основе его построения лежит модульный принцип. В настоящей версии программный комплекс поддерживает формат *CD Audio* с частотой дискретизации 44100 Гц и разрешением 16 бит/отсчет, режимы работы моно или стерео. Он содержит следующие части:

- библиотеку исследуемых ПАМ;

- блок обработки звукового материала, позволяющий выбрать тот или иной звуковой файл из базы (библиотеки) данных, выделить фрагмент для исследования, задавая начальную и конечную точки исследуемого фрагмента отрывка;

- блок статистической обработки и визуализации результатов вычислений, выполняемых в каждой исследуемой ПАМ.

Результаты вычислений, выполняемых в ПАМ, выводятся в форме массива данных, элементами которого являются значения глобального порога маскировки, дБ, отношения сигнал-маска SMR, дБ, а также и число бит  $m$ , требуемых для прозрачного кодирования каждого элемента выборки звукового сигнала для каждой из субполос психоакустического анализа.

Вся информация выводится в отдельный файл электронных таблиц *Microsoft Excel*, что позволяет открыть его в большинстве программ, используемых для статистического анализа результатов эксперимента, а также для визуализации полученных данных. В качестве ПАМ\_1 использовалась модель, созданная *Fabien Petitcolas*. Программные модели ПАМ\_2 и ПАМ\_ААС были созданы ранее аспирантами СПбГУТ им. проф. М.А. Бонч-Бруевича, код вейвлетной ПАМ заимствован из диссертации Г.Г. Рогозинского. Для реализации перечисленных выше функций, а также и интерфейса аппаратно-программного комплекса были написаны соответствующие скрипты. Правильность работы каждой исследуемой ПАМ была предварительно проверена на тональных сигналах разных частот.

В общей сложности была оценена работа каждой модели на 60-ти отрывках звуковых сигналов. В качестве примера на рисунке 1 для ПАМ\_ААС приведена типичная зависимость отношения сигнал-маска ( $a$ ) и число бит ( $b$ ), требуемое для прозрачного кодирования четырех звуковых отрывков.

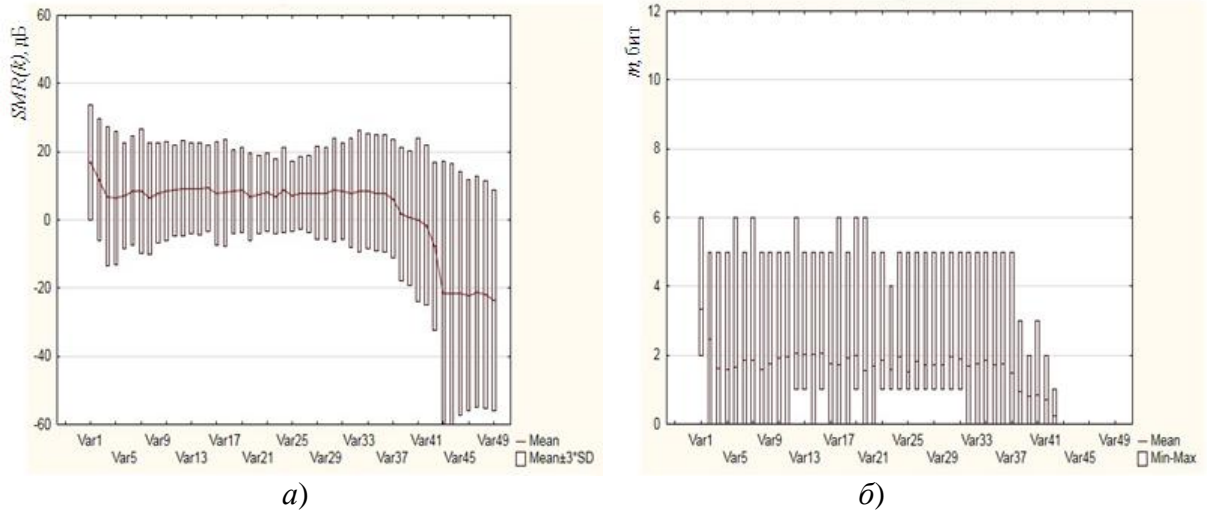


Рисунок 1. Отношение сигнал-маска ( $a$ ) и требуемое ПАМ\_ААС для прозрачного кодирования число бит ( $b$ ); поп-музыка (усредненные данные для четырех отрывков, из которых для исследования взяты по три фрагмента, каждый длиной 15 с). Сплошная линия – среднее значение, верхняя и нижняя границы прямоугольников соответствуют среднему квадратическому отклонению  $\pm 3\sigma$

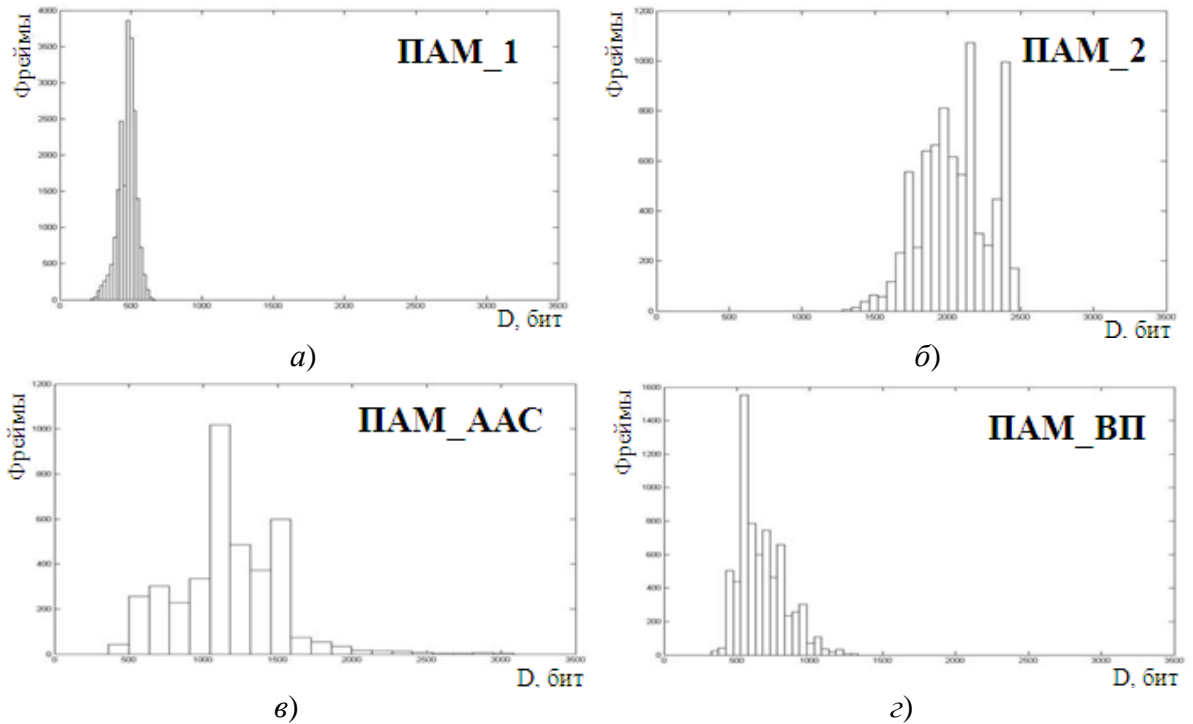


Рисунок 2. Гистограммы распределения бит, требуемых ПАМ для прозрачного кодирования выборки звукового сигнала (рок-музыка): *а* - ПАМ\_1; *б* – ПАМ\_2; *в* – ПАМ\_ААС; *г* – ПАМ\_ВП (вейвлет Добеши 4)

Примеры гистограмм распределения бит (D), требуемых каждой ПАМ для прозрачного кодирования одной выборки звукового сигнала, представлены на рисунке 2 (рок-музыка). Здесь по оси абсцисс отложено число бит, требуемое ПАМ для прозрачного кодирования отсчетов одной выборки звукового сигнала, по оси ординат – число обработанных выборок. Длина выборки звукового сигнала для каждой ПАМ разная.

В результате выполненного исследования ПАМ\_ААС была признана обладающей наилучшими характеристиками при оценке требуемого для прозрачного кодирования числа бит. Однако она разработана для алгоритма компрессии с применением банка полифазных квадратурных зеркальных фильтров (PQMF-фильтров) с очень высокой избирательностью и использованием кодирования коэффициентов модифицированного дискретного косинусного преобразования (МДКП). Для применения ПАМ\_ААС в вейвлетном кодеке, дерево фильтров которого обладает малой избирательностью, требуется проведение дополнительных исследований, что и было проделано в работе. Важно, что в ПАМ\_ААС используются полосы психоакустического анализа разной ширины, их число равно 49, что позволяет более точно учесть особенности распределения энергии звукового сигнала по частоте.

При разработке кодеров работающих в вейвлетном пространстве, необходимо дополнительно изучить особенности распределения энергии шумов квантования по частоте и разработать метод их учета при кодировании коэффициентов вейвлетного дискретного преобразования.

**В третьей главе** излагаются разработанные в рамках диссертации метод применения ПАМ\_ААС к вейвлетному домену и алгоритм контроля энергии шумов квантования в субполосах кодирования вейвлетных коэффициентов.

Вычисление порога маскировки непосредственно в домене вейвлетного преобразования - это более простой алгоритм, позволяющий выполнять все операции в одном



домене с учетом его особенностей. В то же время существует ряд проблем при использовании результатов психоакустического анализа, выполненного непосредственно в вейвлетном домене. Самая главная из этих проблем связана с тем, что дискретное вейвлетное преобразование не обеспечивает компактного распределения по частоте энергии для чистых тонов.

В качестве примера на рисунке 3 показаны ступенчатые функции распределения энергии вейвлетных коэффициентов  $S(i)$ , в дБ, в 28-ми полосах дерева анализирующего фильтра для тона с частотой 10000 Гц. Вейвлетное преобразование исходного сигнала было выполнено с использованием вейвлетов Добеши 4 (сплошная линия) и Добеши 20 (пунктирная линия).

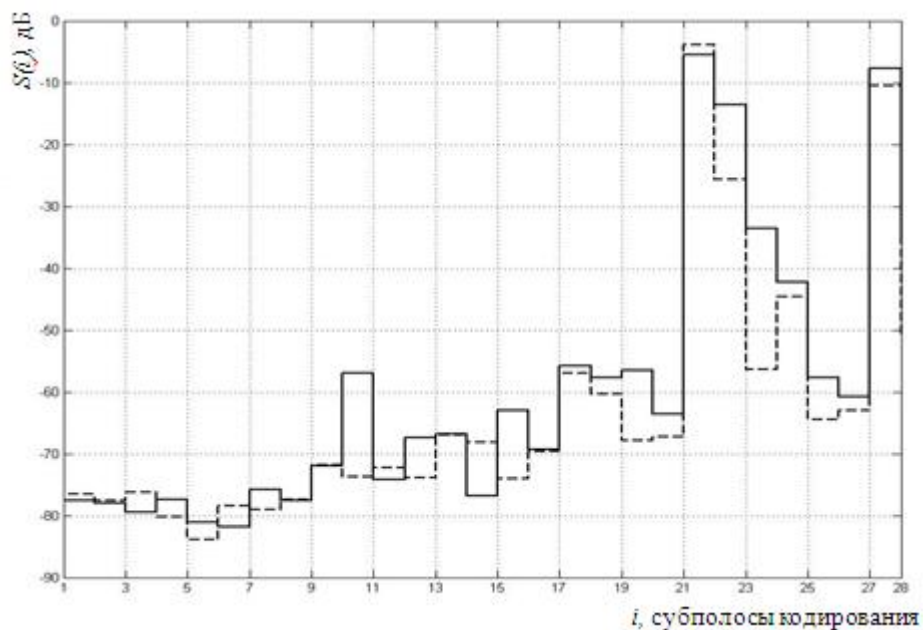


Рисунок 3. Энергетический спектр вейвлетных коэффициентов для тонального сигнала частотой 10000 Гц

Даже вейвлет Добеши 20 (рисунок 3) с длиной импульсной характеристики (40 отсчетов звукового сигнала), лишь незначительно повышает локализацию энергии вейвлетных коэффициентов для тональных сигналов.

Так как вейвлетное преобразование обладает плохой частотной локализацией энергии тональных сигналов в спектре вейвлетных коэффициентов, пороги маскировки, полученные на основе вычисления энергии вейвлетных полос (без учета влияния энергии сигнала в других полосах), как показали проведенные исследования, не соответствуют реальным значениям.

Вследствие недостаточной избирательности анализирующих/синтезирующих QMF-фильтров шумы квантования, появляющиеся на одном уровне вейвлетного преобразования, после восстановления выборки сигнала в декодере будут присутствовать с разным уровнем в различных участках частотного спектра, не только в той полосе кодирования, в которой находится полезный сигнал, как это имеет место в случае применения полифазных квадратурных зеркальных фильтров (PQMF). Следовательно, при расчете порога маскировки в случае с вейвлетами необходимо учитывать не только область частот вблизи маскирующего сигнала, но и другие, несмежные с ней области. Зависимости, представленные на рисунке 4, наглядно показывают, что шумы квантования при обратном дискретном вейвлетном преобразовании в разной степени распространяются по всему частотному спектру (при наличии шума в анализирующем дереве фильт-

ров только в одной полосе кодирования). При этом имеет место сложная по форме огибающая частотного спектра.

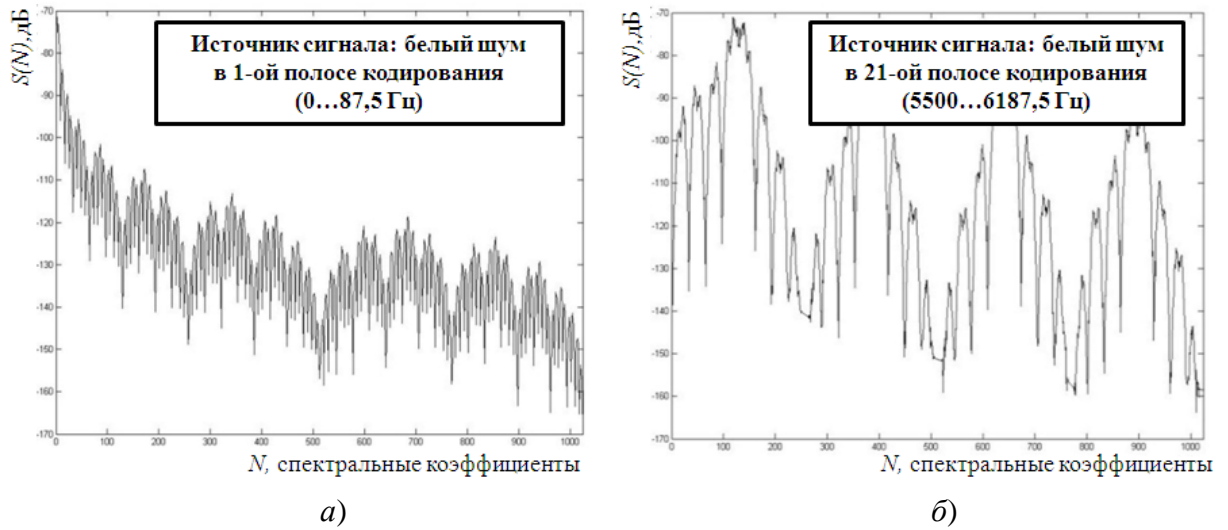


Рисунок 4. Энергетический частотный спектр вейвлетных коэффициентов для шумов квантования при наличии шума: а- в полосе 1, вейвлет Добеши 4; б- в полосе 9, вейвлет Добеши 4

Выходом ПАМ\_ААС является вектор из 49 значений отношения сигнал-маска (в случае с длинным блоком и частотой дискретизации 44100 Гц). На рисунке 5 спектральные линии энергетического спектра восстановленного сигнала сгруппированы в соответствии с правилом, принятым в ПАМ\_ААС (ступенчатая кривая):

$$T(k) = \max \left\{ N_m \left( N_H \right) \right\} \quad (1)$$

где  $T(k)$  – максимальный уровень энергии сигнала в полосе психоакустического анализа  $k$ , дБ,  $N_m(k)$  – номер (индекс) спектральной линии, соответствующей нижней граничной частоте полосы психоакустического анализа  $k$ ,  $N_H(k)$  – номер (индекс) спектральной линии, соответствующей верхней граничной частоте полосы психоакустического анализа  $k$ .

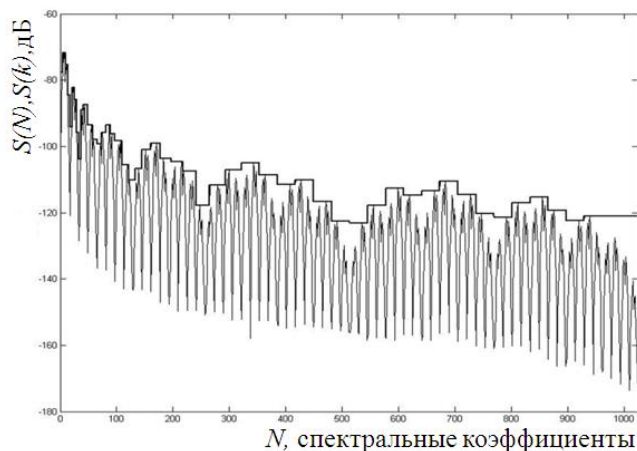


Рисунок 5. Значения энергии шумов квантования, вычисленные для полос психоакустического анализа ПАМ\_ААС

Данный подход позволяет описать в виде вектора, состоящего из 49 элементов, энергетический спектр шумов квантования для каждого терминального узла вейвлетного дерева в отдельности. В данном случае мы имеем матрицу  $\mathbf{M}$  размерностью  $28 \times 49$ , состоящую из 28 строк и 49 столбцов. Элементами матрицы являются значения энергии шумов квантования, выраженные в дБ. При наличии подобной матрицы для максимального уровня энергии шумов квантования остальные значения могут быть получены по правилу:

$$N_m = N_{\max} - 6,02 \cdot m, \text{ дБ}, \quad (2)$$

где  $N_{\max}$  – максимальный уровень энергии шумов квантования (для кодового слова длиной 2 бита),  $m$  – количество бит, используемых для кодирования вейвлетных коэффициентов в рассматриваемой полосе (без учета знакового бита).

Схема, иллюстрирующая расчет энергии шумов квантования для одной строки матрицы  $\mathbf{M}$  для субполосы кодирования 4, приведена на рисунке 6. В начале расчета синтезируется дерево нулевых вейвлетных коэффициентов, т.е. содержащее в качестве коэффициентов вейвлетного преобразования только нулевые значения. К одному из терминальных узлов (на рисунке 6 – узел 4) подключается генератор белого шума, имитирующий шум квантования.

После этого выполняется обратное вейвлетное преобразование, восстанавливающее исходный сигнал во времени  $s[n]$  и вычисляется его амплитудный спектр  $S[k]$ . Далее спектральные линии группируются в соответствии с полосами психоакустического анализа ПАМ\_ААС ( $SF_i$ ). Для каждой полосы анализа ПАМ\_ААС определяется максимальное значение энергии шумов квантования. Это значение сохраняется в качестве соответствующего элемента матрицы  $\mathbf{M}$  уровней энергии шумов квантования. Данная процедура поочередно выполняется для всех 28-ми полос кодирования. Новое значение энергии шумов квантования в каждой полосе психоакустического анализа при этом суммируется с уже полученным ранее значением. При многократном выполнении данной совокупности процедур для каждой из 49 полос психоакустического анализа, наблюдается некоторый разброс максимальных значений энергии шумов квантования. В частности, при выполнении 1000 таких итераций он в среднем составляет около 10 дБ.

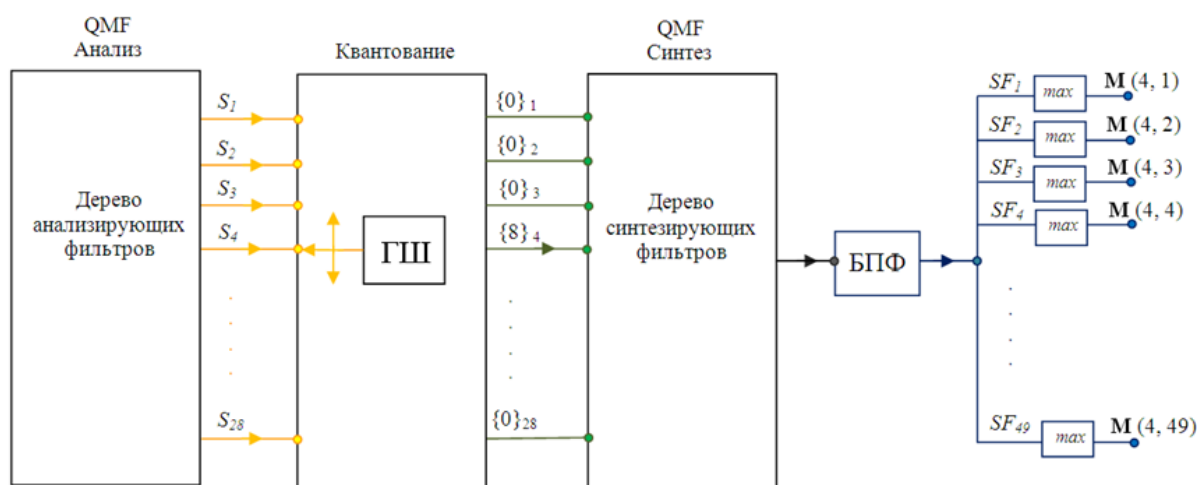


Рисунок 6. Схема вычисления элементов матрицы  $\mathbf{M}$  ( $28 \times 49$ )

В результате такой серии экспериментов при фиксированном типе вейвлетной функции была получена матрица  $\mathbf{M}$  шумов квантования размером 28x49. На рисунке 7, *a* представлена её визуализация. Яркость элементов здесь говорит о величине энергии шума квантования, дБ, в соответствии со шкалой, представленной на рисунке 7, *a* чуть правее от рисунка самой матрицы.

Наличие матрицы энергии шумов квантования вейвлетных коэффициентов позволяет предложить принцип прозрачного кодирования, основанный на контроле допустимого уровня энергии шумов квантования в полосах психоакустического анализа.

В соответствии с принципами психоакустической компрессии для незаметного на слух устранения психоакустической избыточности звукового сигнала необходимо, чтобы значение энергии шумов квантования в соответствующей полосе кодирования оставалось ниже уровня маски, рассчитанной психоакустической моделью ПАМ\_ААС.

Предлагаемый в работе алгоритм контроля уровня энергии шумов квантования приведен на рисунке 7, *б*, где:  $i$  - индекс (номер) субполосы кодирования;  $k$  – индекс (номер) полосы психоакустического анализа ПАМ\_ААС;  $T(k)$  – значение порога маскировки, дБ для полосы психоакустического анализа  $k$ ;  $N_i(k)$  – фактическое значение уровня энергии шумов квантования, дБ в полосе психоакустического анализа  $k$ . Значение  $B(k)$  вектора шагов масштабирования уровня энергии шумов квантования для  $k$ -ой полосы определяет минимальное количество битов кодовых слов вейвлетных коэффициентов для данной полосы кодирования.

Поясним кратко суть данного алгоритма (рисунок 7, *б*). Рассматривается энергетический спектр источника шумов квантования  $i$ -ой полосы ( $i$ -ая строка матрицы  $\mathbf{M}$ ; рисунок 8, пунктирная кривая). Значения вектора уровня энергии шума квантования  $N_i$  в полосе кодирования  $i$  подвергаются масштабированию с шагом 6,02 дБ до тех пор, пока они не станут ниже порога маскировки  $T(k)$  (сплошная линия на рисунке 8) в каждой из 49 полос психоакустического анализа.

Как только будет достигнуто максимально допустимое значение уровня энергии шумов квантования в полосе кодирования  $i$ , выбирается следующая строка матрицы  $\mathbf{M}$ . Её значения перед проверкой на превышение порога маскировки поэлементно суммируются с рассчитанными ранее значениями энергии шумов квантования вейвлетных источников шума в каждой полосе. В результате этих итераций учитывается вклад шумов квантования вейвлетных коэффициентов каждой из 28 полос.

Процедура контроля уровня шумов квантования происходит итеративно до тех пор, пока не будет достигнуто наилучшее распределение их допустимого значения вблизи порога маскировки. Количество шагов  $L$  масштабирования уровня энергии шумов квантования  $k$ -ой полосы определяет минимальное количество битов кодовых слов вейвлетных коэффициентов для данной полосы кодирования (без учета знакового бита).

Дополнительное увеличение коэффициента сжатия может быть достигнуто при обнулении вейвлетных коэффициентов уровней декомпозиции с малыми значениями энергии, однако это требует проведения дополнительных исследований.

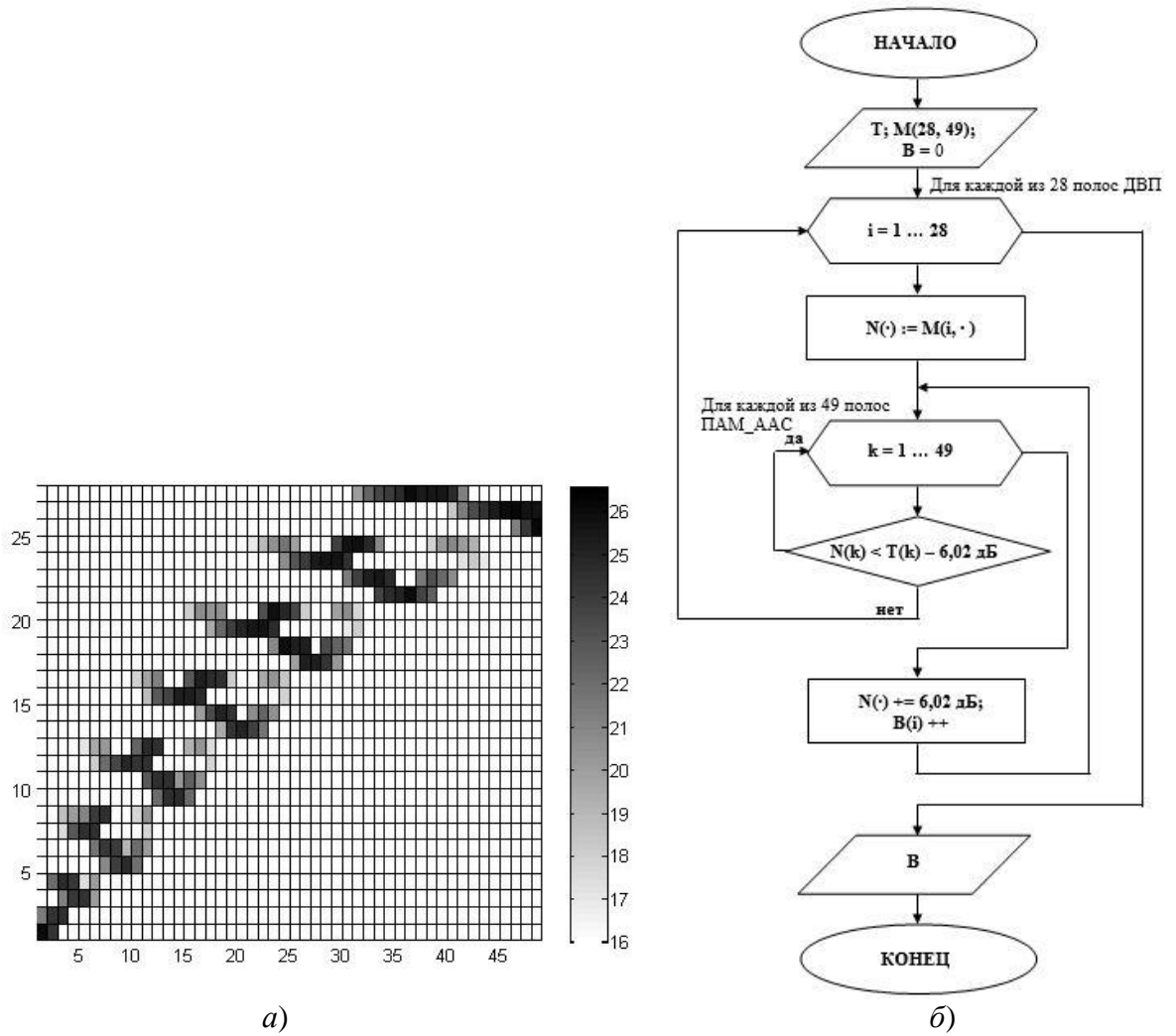


Рисунок 7. Визуализация матрицы  $M$  (28 x 49) шумов квантования (а) и схема алгоритма контроля уровней энергии шумов квантования (б)



Рисунок 8. Пример контроля уровня энергии шумов квантования.

**Четвертая глава** содержит структурную схему разработанного экспериментального кодека, описание основных его программных модулей представленных на рисунке 9 и результаты экспертной оценки качества компрессии цифровых аудиоданных.

Исходная выборка звукового сигнала длиной 2048 ИКМ-отсчетов поступает на вход блока ПДВП. Эти же отсчеты поступают параллельно в блок ПАМ\_ААС. При этом используется техника overlap-and-add, при которой реализуется 50% перекрытие выборок с оконной функцией, необходимой для уменьшения искажений, вызванных ортогональным преобразованием.

Пакетное дискретное вейвлетное преобразование (ПДВП) выполняется с выбранным типом вейвлета. На выходе блока ПДВП для базового профиля мы имеем 28 групп вейвлетных коэффициентов различной длины – от 4-х коэффициентов для полос кодирования 1-8 и до 256-ти коэффициентов для полос кодирования с номерами 25-28.

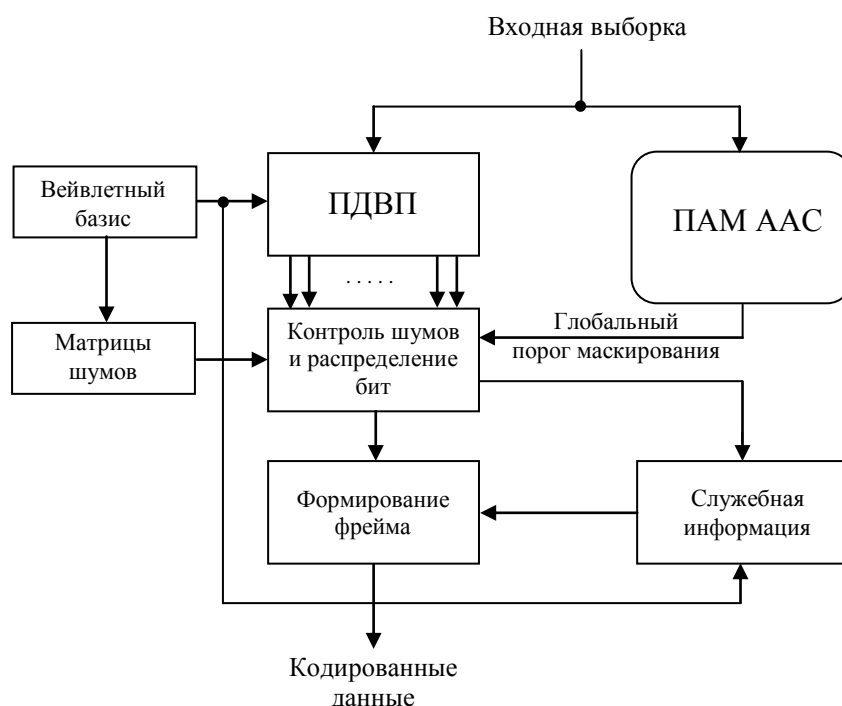


Рисунок 9. Структурная схема вейвлетного кодера. Базовый профиль

Коэффициенты прямого дискретного вейвлетного преобразования подвергаются повторному квантованию в соответствии с результатами психоакустического анализа. В процессе создания программной модели были приняты некоторые допущения с целью уменьшения количества выполняемых вычислительных операций и исключения возможности возникновения ошибок в тех частях модели, которые не имеют непосредственного отношения к предмету исследования. Среди этих допущений следует отметить:

- кодер и декодер объединены в общую программу. Выходной файл в формате сжатых данных не создается. Фактически, сразу после повторного квантования вейвлетных коэффициентов, соответствующих временным отсчетам текущей выборки, происходит их восстановление. Тем самым исключаются из моделирования такие однозначные процессы, как формирование аудиофрейма, запись сжатого формата данных, его считывание, декодирование заголовков и т.д.;

- таблица масштабных коэффициентов не используется. За масштабный коэффициент принимается максимальное по модулю значение вектора вейвлетных коэффици-

ентов  $i$ -ой полосы кодирования. Таким образом, в каждой полосе кодирования после масштабирования присутствует значение вейвлетного коэффициента равное по модулю 1;

- программная модель поддерживает в качестве входных данных звуковые файлы формата *wav* с частотой дискретизации 44100 Гц, моно или стерео;

- скорость цифрового потока для кодера не задается. В результате кодирования выбирается минимальное количество битов, при котором сжатие цифровых аудиоданных остается прозрачным (шумы квантования в каждой полосе кодирования остаются ниже или лежат на пороге слышимости).

Исследования показали, что существует определенный запас  $\Delta$  по уровню между порогом маскировки и допустимым уровнем шумов квантования, соответствующий примерно одному-двум битам. Иными словами, длина кодовых слов при повторном кодировании коэффициентов вейвлетного преобразования в каждой из субполос может быть уменьшена дополнительно ещё на 1...2 разряда.

#### **Экспертная оценка качества работы вейвлетного кодека.**

Экспертная оценка качества работы кодека применялась для:

- а) экспериментального подтверждения прозрачности вейвлетной компрессии при минимальном количестве выделенных бит;

- б) экспериментального подтверждения наличия некоторого запаса  $\Delta$  при расчете глобального порога маскировки в каждой из субполос кодирования.

Экспертная оценка качества работы кодека была выполнена в соответствии с рекомендацией ITU-R BS.1116.

При проведении экспертиз использовалось две группы слушателей. Одна группа была сформирована из студентов-звукорежиссеров Санкт-Петербургского государственного института кино и телевидения, другая – из студентов Санкт-Петербургского государственного университета телекоммуникаций им. проф. М.А. Бонч-Бруевича, не имеющих специальной подготовки. В каждую группу слушателей входило 12 человек (по шесть представителей каждого пола) в возрасте от 17 до 23 лет.

Для всех тестовых отрывков, выполненное кодером минимальное распределение битов, обеспечило прозрачное кодирование. Дополнительное уменьшение количества выделяемых бит на один разряд ( $\Delta=1$ ) при кодировании вейвлетных коэффициентов в каждой полосе так же осталось незаметным для большинства слушателей. Оценки различия сравниваемых сигналов составляли после обработки 4,9-5,0 балла (звучания равноценны).

Уменьшение количества выделяемых битов на два разряда ( $\Delta=2$ ) в каждой субполосе кодирования привело к возникновению незначительных артефактов в виде дроблений и модуляций звука. Оценки различия звучаний в паре находились в пределах 3,7-4,5 балла.

При уменьшении длины кодовых слов на 3 разряда ( $\Delta=3$ ) восстановленный материал содержал артефакты, заметные подавляющему большинству экспертов.

Результаты экспериментов подтвердили корректность работы алгоритма и адекватность применения ПАМ\_ААС к кодеру, работающему в вейвлетном пространстве. Образовавшийся запас по уровню энергии шумов квантования может быть компенсирован вводом коэффициента запаса  $\Delta=1$  в алгоритм распределения бит.

В таблице 1 и 2 приведены результаты статистического анализа выделения модели бит на кодирование отчетов одной выборки сигнала для тестовых отрывков различных жанров при использовании преобразования с вейвлетом Добеши 10 для значений  $\Delta=0$  и  $\Delta=1$ .

Таблица 1.

Вейвлет Добеши 10 ( $\Delta=0$ )

Жанр	Среднее	Min	Max	СКО
Хоровая	8521	464	12976	1902
Танцевальная	11139	296	14056	1457
Камерная	8073	0	12600	2513
Оркестровая	9449	4496	12680	1565
Поп-музыка	10895	4064	14024	1411
Рок-музыка	10567	3056	13688	1722

Таблица 2.

Вейвлет Добеши 10 ( $\Delta=1$ )

Жанр	Среднее	Min	Max	СКО
Хоровая	5358	336	10416	1466
Танцевальная	8562	160	11976	1492
Камерная	4876	0	9560	1942
Оркестровая	6374	2160	10120	1143
Поп-музыка	8230	3224	11976	1316
Рок-музыка	7827	2160	11640	1772

Кодек, работающий в вейвлетном пространстве, обеспечивает прозрачное кодирование при средней длине кодовых слов вейвлетных коэффициентов 3,35...4,7 бит при использовании вейвлетного базиса Добеши 10 и 3,71...5,22 бит при использовании вейвлетного базиса Добеши 4.

Переход к вейвлету более высокого порядка (от Добеши 4 к Добеши 10) обеспечивает уменьшение количества бит, выделяемых на фрейм без потери качества звучания, на 4 – 13 %. При этом наилучшие показатели обеспечиваются для органно-хоровой и камерной музыки. Наихудшие показатели и в этом случае были у жанров танцевальной музыки, рока и поп-музыки.

Ввод поправочного коэффициента  $\Delta=1$  обеспечил уменьшение количества бит, выделяемых на кодирование выборки звукового сигнала без потери качества звучания, на 23 – 39 % в случае преобразования с вейвлетом Добеши 10, и на 22 – 36 % в случае преобразования с вейвлетом Добеши 4.

#### **Выводы.**

1. Проведён сравнительный анализ четырех психоакустических моделей, используемых в современных алгоритмах компрессии цифровых аудиоданных. Показано, что из всех рассмотренных моделей, психоакустическая модель кодера стандарта MPEG-2 ISO/IEC 13818-7 AAC обладает наилучшим частотным разрешением, является более гибкой, чем другие модели, выделяет в среднем, меньший процент бит от их максимального количества, доступных при установленной скорости цифрового потока. Однако для применения её в вейвлетном кодеке требуется проведение дополнительных исследований.

2. Особенности дискретного вейвлетного преобразования, связанные с недостаточно высокой частотной избирательностью квадратурно-зеркальных фильтров (QMF-фильтров), существенно ухудшают частотную локализацию энергии обрабатываемого сигнала выборки. Этот феномен наиболее заметен на гармонических сигналах, энергетический спектр которых при вейвлетном преобразовании, как правило, содержит несколько энергетических максимумов, расположенных в различных несмежных частях спектра.

3. Порог маскировки, полученный по вектору энергий вейвлетных коэффициентов только в одной субполосе кодирования (как это имеет место в известных алгоритмах компрессии, например, группы MPEG) без учета влияния других субполос, дает неадек-



ватную слуховому восприятию оценку данного порога. Вследствие этого он не может быть использован при оценке требуемой минимальной длины кодовых слов квантованных значений коэффициентов прямого дискретного вейвлетного преобразования, требуемых для прозрачного кодирования.

**4.** В работе предложен, обоснован и экспериментально проверен метод исследования, позволяющий:

- изучить особенности распределения энергии шумов квантования вейвлетных коэффициентов по частоте;

- рассчитать энергию искажений квантования для каждого терминального узла дерева вейвлетного анализа/синтеза;

- представить полученные результаты в виде матрицы шумов квантования **M** (28x49), содержащей 28 строк по числу субполос анализирующего фильтра и 49 столбцов по числу полос психоакустического анализа ПАМ\_ААС, что необходимо и достаточно для последующего расчета числа бит, требуемых для прозрачного кодирования вейвлетных коэффициентов в каждой из субполос кодирования.

**5.** Разработан метод контроля энергии шумов квантования, позволяющий, используя матрицу **M**, элементами которой являются значения энергии шумов квантования, психоакустическую модель ПАМ\_ААС и предложенный в работе итерационный алгоритм масштабирования энергии шумов квантования, найти требуемую для прозрачного кодирования длину кодовых слов вейвлетных коэффициентов в каждой из субполос кодирования.

**6.** Предложено два алгоритма сжатия звука с потерями на основе вейвлетного преобразования с использованием ПАМ\_ААС и алгоритма контроля уровня шумов квантования, отличающихся различной степенью сложности. Первый алгоритм работает с фиксированным деревом вейвлетной декомпозиции и фиксированным типом вейвлета. Второй алгоритм повышенной сложности позволяет в результате предварительного анализа сигнала выборки звукового сигнала подобрать для него наилучший вейвлетный базис и адаптировать с учетом свойств сигнала выборки дерево фильтров самой декомпозиции. В работе первый алгоритм компрессии цифровых аудиоданных реализован программно в среде MATLAB.

**7.** Данные субъективно-статистических экспертиз подтвердили, что при использовании вейвлета Добеши 10 на кодирование одной выборки длиной 2048 отсчетов музыкальных отрывков вейвлетный кодек тратит в среднем:

- при  $\Delta=0$  примерно 9774 бит (средняя длина кодовых слов вейвлетных коэффициентов составляет здесь 4,77 бит);

- при  $\Delta=1$  (где также еще практически обеспечивается прозрачность кодирования) эта величина составляет соответственно 6868 бит при средней длине кодовых слов вейвлетных коэффициентов равной 3,35 бит.

При применении в кодеке вейвлета Добеши 4 прозрачное кодирование музыкальных фрагментов одной выборки длиной 2048 отсчетов требует в среднем 10774 бит при среднем значении длины кодового слова 5,22 бит ( $\Delta=0$ ) и соответственно 7604 бит при средней длине кодового слова вейвлетных коэффициентов равной 3,71 бит ( $\Delta=1$ ).

**8.** Переход к вейвету более высокого порядка (от Добеши 4 к Добеши 10) обеспечивает уменьшение количества бит, выделяемых на одну выборку длиной 2048 отсчетов без потери качества звучания, на 4...13 %. Наилучшие показатели обеспечиваются для органно-хоровой и камерной музыки. Это связано с характером этих жанров. Вейветы более высокого порядка обеспечивают лучшую частотную локализацию. Наихудшие показатели у танцевальной музыки, рока и поп-музыки – жанров, в статистическом отношении находящихся ближе к окрашенному шуму.

Ввод поправочного коэффициента  $\Delta=1$  обеспечил уменьшение количества бит, выделяемых на одну выборку длиной 2048 отсчетов без потери качества звучания, на 23...39 % в случае преобразования с вейвлетом Добеши 10, и на 22...36 % в случае преобразования с вейвлетом Добеши 4.

9. Дополнительное уменьшение скорости цифрового потока может обеспечить обнуление вейвлетных коэффициентов некоторых уровней декомпозиции дерева вейвлетного анализа. Рассмотренный в работе подход позволяет и в этом случае оценить величину энергии искажений квантования и принять решение о целесообразности сохранения того или иного уровня энергии вейвлетных коэффициентов.

10. На основе выполненной работы разработан лабораторный учебный комплекс на базе MATLAB, позволяющий подключать в виде модулей различные психоакустические модели, исследовать характер распределения бит по субполосам кодирования, оценивать пороги маскирования и т.п., а также автоматизировать процесс обработки результатов экспериментальных исследований.

### СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРАЦИИ

1. Фадеев Д.Р. Оценка качества и эффективности работы кодека в вейвлетном пространстве / Ковалгин Ю.А., Фадеев Д.Р. // Системы управления и информационные технологии, Научно-технический журнал, №2 (64), 2016 – С. 58-62.

2. Фадеев Д.Р. Исследование психоакустических моделей кодеков с компрессией цифровых аудиоданных / Ковалгин Ю.А., Фадеев Д.Р. // Современная наука актуальные проблемы теории и практики, Серия «Естественные и технические науки», 2016, №7, 2016. – С. 29-38.

3. Фадеев Д.Р. Контроль искажений квантования при работе кодека в вейвлетном пространстве / Ковалгин Ю.А., Фадеев Д.Р. // Электросвязь, Тема номера « Интернет вещей и цифровая экономика», № 9, 2016. – С. 68-73.

4. Фадеев Д.Р. Программный комплекс для исследования эффективности психоакустических моделей в алгоритмах перцепционного сжатия звуковых сигналов / Фадеев Д.Р. // Системы управления и информационные технологии, Научно-технический журнал, №1(63), 2016. – С. 47-50.

5. Фадеев Д.Р. Исследование эффективности психоакустических моделей в алгоритмах перцепционного сжатия звуковых сигналов / Фадеев Д.Р. // Приоритетные направления развития науки и технологий. Сборник докладов XVIII международной научно-технической конференции, издательство «Инновационные технологии», Тула, 2015. – С. 70-72.

6. Фадеев Д.Р. Вейвлетная психоакустическая модель с переключением деревьев / Фадеев Д.Р. // Приоритетные направления развития науки и технологий. Сборник докладов XVIII международной научно-технической конференции, издательство «Инновационные технологии», Тула, 2015 – С. 72-75.

7. Фадеев Д.Р. Применение вейвлет-преобразований в алгоритмах компрессии цифровых аудиоданных / Фадеев Д.Р. // Сборник: Актуальные проблемы инфотелекоммуникаций в науке и образовании «международная научно-техническая научно-методическая конференция» 2013. – С. 357-363.

8. Фадеев Д.Р. Метод контроля шумов квантования вейвлетных коэффициентов в перцепционном сжатии звуковых сигналов / Фадеев Д.Р. // Актуальные проблемы инфотелекоммуникаций в науке и образовании»; СПбГУТ.-СПб., 2016.

9. Фадеев Д.Р. Учет искажений квантования при кодировании коэффициентов вейвлетного преобразования / Фадеев Д.Р. // Актуальные проблемы радио- и кинотехнологий: материалы Международной научно-технической конференции, посвященной 125-летию со дня рождения русского радиотехника, проф. Ленинградского института киноинженеров В.И. Волынкина.: СПбГИКиТ, 2016. – С. 48.