

На правах рукописи



Шиленков Егор Андреевич

**Структурно-параметрический синтез ортогонального речевого кодера
адаптивного критериям степени сжатия и разборчивости речи**

Специальность: 05.13.01 – Системный анализ, управление и обработка
информации (технические системы)

Автореферат
диссертации на соискание учёной степени
кандидата технических наук

Курск – 2016

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего образования «Юго–Западный государственный университет»

Научный руководитель: кандидат технических наук, профессор
Потапенко Александр Михайлович,
Юго-Западный государственный университет,
профессор кафедры космического приборостроения и систем связи

Официальные оппоненты: Сарычев Валентин Александрович,
доктор технических наук, профессор,
Акционерное общество «Научно-производственное предприятие «Радар ммс»,
заместитель генерального конструктора;

Корнеев Юрий Алексеевич,
кандидат технических наук,
ФГАОУ ВО «Санкт-Петербургский государственный университет аэрокосмического приборостроения», доцент «Института радиотехники, электроники и связи (Институт №2 ГУАП)»,
старший научный сотрудник.

Ведущая организация ФГАОУ ВО «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики» (Университет ИТМО), г. Санкт-Петербург.

Защита состоится 24 ноября 2016 года в 16:00 на заседании диссертационного совета Д 002.199.01 при Федеральном государственном бюджетном учреждении науки Санкт-Петербургском институте информатики и автоматизации Российской академии наук (СПИИРАН) по адресу: 199178, Санкт-Петербург, В.О., 14 линия, 39.

С диссертацией можно ознакомиться в библиотеке и на сайте Федерального государственного бюджетного учреждения науки Санкт-Петербургского института информатики и автоматизации Российской академии наук.

Автореферат разослан _____ 20__ года.

Учёный секретарь
диссертационного совета Д 002.199.01,
канд. техн. наук, доцент

Фаткиева Роза Равильевна

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. В процессах сбора, обмена и обработки информации одно из важных мест занимают вопросы её компактного описания. Это подтверждается ростом вида и числа оборудования преобразования и хранения данных, используемого в локальных и глобальных системах селекции, передачи и представления информации. При этом в ряде важных практических приложений объективно имеет место противоречие между потребностью повышения степени компактности (сжатия) описания исходных данных и необходимостью сохранения требуемого уровня их качества. В этой связи проблема повышения степени сжатия описания данных в системах сбора, обработки, хранения и представления является актуальной. Во всей совокупности задач, направленных на разрешение данных проблем, значительное место занимают задачи сжатия аудиоданных. И это связано с широким разнообразием целей и условий аудиообмена и, соответственно, разнообразием требований к степени сжатия и допустимому уровню качества воспроизведения речевых данных.

Степень разработанности темы. Вопросам сжатия данных в общем, и речевых в частности, начиная со второй половины XX века, постоянно уделялось должное внимание российской и зарубежной наукой. В трудах классиков Котельникова В.А., Давыдова А.В., Шпигеля М.Ф., Бернштейна Л.Р. обоснованы основополагающие вопросы цифрового преобразования речевых сигналов. В работах Сергиенко А.Б., Борискевича А.А., Ронжина А.Л., Алдошиной И.А., Склера Б., Кестера У., Терхарда Э., Янга К. детально рассмотрены методы и способы реализации процедур сжатия звукового потока данных. Полученные в этих и других работах результаты создают необходимую научно-методическую и практическую базу для дальнейшего повышения эффективности процессов хранения и передачи аудиоданных в системах их сбора, обработки и представления.

В качестве одного из важных перспективных направлений исследований представляются такие как: более полный учёт нестационарности свойств исходных речевых сигналов и условий их формирования, разработка методов речевых кодеков с динамической структурой, алгоритмической и параметрической адаптацией. Достижение приемлемых результатов возможно, в основном, за счёт сжатия потока данных канального уровня в устройствах хранения буферных данных. Таким образом, преимущество достигается теми устройствами связи, в которых показатель качества предоставляемой услуги связи балансирует с пониженными скоростями передачи данных.

Начиная со второй половины XX века, с развитием цифровых технологий передачи данных параллельно предпринимались подходы к техникам оцифровки речевых аудио-сигналов. Первые ИКМ кодеры применяли технологии неравномерного квантования для снижения разрядности кода не в ущерб качеству речи. Далее были замечены особенности произношения слов и фраз человеком, связанные с тем, что уровень громкости голоса в динамике скачкообразно измениться не может и описывается известными математическими функциями. Данная особенность была реализована в кодере с адаптивным предсказанием, который успешно применяется и в настоящее время в сотовой, интернет связи и т.д. В случае с большим количеством передаваемой информации получили развитие ортогональ-

ные кодеры, в которых реализована особенность неравномерного частотного восприятия информации человеком.

Однако главными требованиями к речевому общению являются не только критерии разборчивости, но и задержка сигнала. Так, в области критерия степени сжатия лидируют ортогональные кодеры. В тоже время минимальная задержка является главным преимуществом адаптивных кодеров с линейным предсказанием. В настоящее время ведутся активные исследования по поиску гибридных решений.

Объектом исследования выбраны устройства цифровой обработки звуковых сигналов в системах их хранения и передачи.

Предметом работы являются алгоритмы и процедуры сжатия потока речевых данных в реальном времени: психоакустическое маскирование, дискретное преобразование, кодирование, цифровая фильтрация, словарная компрессия и предсказание.

Цель и задачи исследования. Целью является повышение эффективности хранения и передачи речевой аудиоинформации в условиях высокой степени нестационарности свойств сигнала. Под повышением эффективности понимается уменьшение объёма аудиоданных при поддержании заданного качества воспроизведения аудиоинформации.

Научно–техническая задача состоит в разработке, исследовании и адаптации методов и алгоритмов кодирования звука, позволяющих оперативно согласовать параметры процедур сжатия нестационарного речевого сигнала для сохранения уровня разборчивости голоса при понижении скорости потока данных.

Использование данных методов в системах передачи, хранения и обработки информации позволит повысить степень их использования при заданных требованиях к её качеству. Таким образом, решается актуальная проблема ухудшения воспроизведения речевого сигнала при уменьшении его объёма.

Для достижения поставленной цели в работе сформулированы задачи.

1. Анализ текущей ситуации в области речевых кодеров для исследования перспективных структур и техник сжатия потока данных.

2. Анализ и исследование методов и процедур преобразования широкополосного звукового сигнала с целью определения наиболее значимых операций для повышения степени его сжатия.

3. Разработка математической модели процессов цифровой обработки и преобразования, обеспечивающей оперативную алгоритмическую и параметрическую адаптивность устройств кодирования и сжатия при высокой степени нестационарности свойств речевого сигнала.

4. Разработка и модификация методик и алгоритмов процедур сжатия в устройствах хранения и передачи сигнала, позволяющих сократить объем данных.

5. Синтез структурно–функциональной схемы речевого кодера, приспособленного к поставленным требованиям передачи речевой информации в реальном времени, и экспериментальная оценка результата сжатия данных.

Научная новизна состоит в следующем:

- разработана математическая модель, основанная на учёте нестационарности свойств речевого сигнала и особенностей его слухового восприятия, позво-

ляющая провести многомерный анализ информации для выбора процедур сжатия и контроля потерь качества;

- реализованы алгоритмы прореживания речевого сигнала, основанные на математическом анализе его свойств, осуществляющие частотно-временные преобразования для уменьшения скорости передачи;

- предложена структурно–функциональная организация и программная реализация речевого кодера, отличающегося выбором перечня алгоритмов и методик сжатия и позволяющего оценить уровень качества воспроизведения речи при изменении объёма информации.

Теоретическая и практическая значимость работы. Теоретическая значимость определена как совокупность разработанных моделей и методик линейной и ортогональной цифровой обработки речевых сигналов:

- модель расчёта коэффициентов импульсной характеристики КИХ–фильтров банка полифазных фильтров, позволяющая управлять критерием выразительности речи диктора;

- математическая модель дискретных частотно-временных преобразований, осуществляющая анализ нестационарного сигнала, для управления критерием разборчивости речи;

- алгоритм предиктора позволяет выявлять невыразительные временные диапазоны речевого сигнала и выполнять сжатие путём прореживания, тем самым используя особенность восприятия нередуцированных звуков русской речи.

Практическая значимость:

- методика и алгоритм построения частотно–временной маски, реализующие прореживание при маскировании слуха, применение которых позволяет сжимать размер дискретных значений до 8 раз при условии разборчивости голоса;

- адаптированные методики сжатия без потерь: словарное сжатие по динамическому и статическому словарю кодовых деревьев; сжатие при помощи скользящего окна, использование которых в совокупности с прореживанием позволяет достигнуть степени сжатия в 14 раз;

- структурный синтез речевого кодера позволяет реализовать средство сжатия голосового сигнала, контролирующее качество воспроизведения речи и скорость потока данных, зависящих от выбранных процедур частотно-временных преобразований;

- методики сжатия могут быть использованы в учебном процессе по дисциплине «Цифровая обработка сигналов».

Работа выполнена по плану инициативных НИР 2011–2016 гг. кафедры Космического приборостроения и систем связи Юго-западного государственного университета:

- разработка материалов в пояснительную записку технического проекта инновационно–образовательного профиля в части разработки технологической и методической поддержки организации инновационного и образовательного процесса на базе использования данных ДЗЗ и ГИС–технологий. Разработка математических моделей в интересах модернизации и создания базовых элементов РНИС, заказчик - «НПК» РЕКОД», г. Москва;

– создание системы дистанционного обучения и учебно–тренировочного модуля в интересах организации подготовки и повышения квалификации специалистов в области использования РКД, заказчик - Администрация Курской области;

– исследование и разработка научно–технических путей создания инфологической системы информационно–аналитического обеспечения научных исследований ВУЗа, заказчик – РФФИ.

Результаты работы подтверждены актами о внедрении.

Методология и методы исследований. Для решения задач в работе использованы методы математического моделирования, теории вероятности и математической статистики, ортогональных преобразований, оконного преобразования, скользящего окна, цифровой обработки сигналов.

Для анализа входного и выходного сигнала проведена разработка программного обеспечения имплементации речевого кодека с использованием распределённых сред и параллельного объектно-ориентированного программирования.

Экспериментальная оценка производилась с использованием специально синтезированного модуля, разработанного по принципу модельно-ориентированного проектирования и реализующего работу модели процедур сжатия, их анализ и регулирование путём построения спектров, временных диаграмм и аппаратного воспроизведения речи.

Положения, выносимые на защиту:

1. Математическая модель дискретных преобразований во временной и частотной области, реализующая анализ нестационарного сигнала.

2. Алгоритмы прореживания с помощью наложения частотно-временной маски и редуцированных предсказаний линейного предиктора;

3. Модифицированные методики сжатия ортогонального речевого потока без потерь.

4. Динамическая структура кодера как многофункциональный инструмент для разработки средства сжатия речевого потока и оценки качества воспроизведения.

Степень достоверности и апробация результатов. Достоверность научных результатов подтверждается:

– успешной программной имплементацией ортогонального речевого кодера;
– корректностью построения алгоритмов и выводов математических выражений для построения кодера;

– актами об успешном практическом применении результатов диссертации;
– полученные научные результаты прошли проверку на соответствие требованиям хранения и передачи информации в реальном времени математическим моделированием.

Отдельные вопросы были обсуждены на научно–технических конференциях и круглых столах: «Актуальные проблемы телекоммуникаций» (г. Курск, 24 апреля 2011г.); «Общество, современная наука и образование: проблемы и перспективы. Часть 10» (г. Тамбов, 30 ноября 2012г.); «Актуальные проблемы инфокоммуникаций» (г. Курск, апрель 2012г., 2013г., 2014г.); «Проектирование и технология

электронных средств» (г. Владимир, ноябрь 2012г., декабрь 2013г.); «Роботизация вооружённых сил российской федерации» (ГНИИЦ РТ МО РФ, февраль 2016г.); «Пути повышения автономности робототехнических комплексов» (II Международный военно-технический форум «АРМИЯ-2016», 8 сентября 2016г.).

Основные результаты внедрены в учебный процесс кафедры Космического приборостроения и систем связи при преподавании дисциплин «Цифровая обработка сигналов и сигнальные процессоры в средствах подвижной радиосвязи», «Техника микропроцессорных средств в коммутации», «Основы построения инфокоммуникационных систем и сетей», «Сети и системы передачи информации».

Публикации. По материалам диссертации были сделаны следующие публикации: опубликовано 20 научных работ, из них 7 статей в реферированных журналах ВАК, 13 докладов на научно–технических конференциях.

Личный вклад автора. Модификация и адаптация процедур сжатия с потерями и без, организация экспериментальных исследований и формулирование основных научных положений сделаны автором самостоятельно.

Структура и объем работы. Диссертация содержит введение, 4 главы, заключение, список использованных источников, включающий 133 наименования и одно приложение. Основная часть изложена на 118 страницах. Общий объём работы – 137 страниц, включая 54 рисунка и 29 таблиц.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении показана история развития техник цифрового кодирования и сжатия речи, обоснована актуальность темы, определена цель, для её достижения поставлены задачи, перечислены методы исследований. Сформулированы научная новизна, научно–техническая задача, практическая значимость и положения, выносимые на защиту.

В первой главе проведены исследования технологий цифрового речевого кодирования. Определена актуальность обеспечения вариативностью средства потокового сжатия речи. Параметрический анализ методов кодирования и алгоритмов сжатия звука выявил критичные для качества воспроизведения речи аспекты. Аналоговое сжатие на стадии квантования сигнала позволит повысить эффективность использования динамического диапазона и является начальным этапом работы кодера. В качестве ключевых параметров определены функции битового распределения, разрядность квантования и частота дискретизации. Установлена взаимосвязь элементов адаптивного дифференцированного взаимодействия, включающих в себя алгоритмы предсказания, в совокупности осуществляющего сжатие временного потока без потери качества воспроизведения. Данная взаимосвязь показала возможность увеличения степени компрессии за счёт повышения скорости адаптации и количества аналитических блоков. Психоакустические ограничения слуха человека позволяют добиться большей эффективности сжатия путём установления неслышимого порога для звуков. Здесь выделены четыре критерия для перцептивного сжатия. Это модели временного и частотного маскирования, характеристика абсолютного порога слуха и параметры критичных диапазонов слуха.

Учитывая перечисленные результаты первичных исследований, была сформулирована расширенная задача структурно-параметрической реализации устройства для анализа эффективности сжатия и воспроизведения.

Во второй главе проведено математическое и методическое приспособление средств компрессии информации. В ходе проведённой адаптации были сформированы предварительные для данного исследования результаты цифрового представления и обработки звука в сочетании с особенностями его восприятия человеком для проведения процедур сжатия. Модификация известных ортогональных методик преобразования временной функции в частоту позволила реализовать инструмент анализа речевого сигнала для выявления избыточности в условиях нестационарности его восприятия. Здесь адаптированный для речи алгоритм БПФ (1) используется как средство обнаружения громких тонов критичных диапазонов (рисунок 1 слева), а применение модифицированного синусно-косинусного преобразования (2,3) ускоряет процесс вычисления, уменьшив размерность вектора дискретных частот (рисунок 1 справа).

$$k(i) = k_0(i) + W_N^k \cdot k_1(i), \quad i = 1 \cdots \frac{N}{4} - 1; \quad k(i + \frac{N}{4}) = k_0(i) - W_N^k \cdot k_1(i), \quad i = 1 \cdots \frac{N}{4} - 1;$$

$$W_N^k = e^{-j \frac{2\pi k}{N}} \quad (1)$$

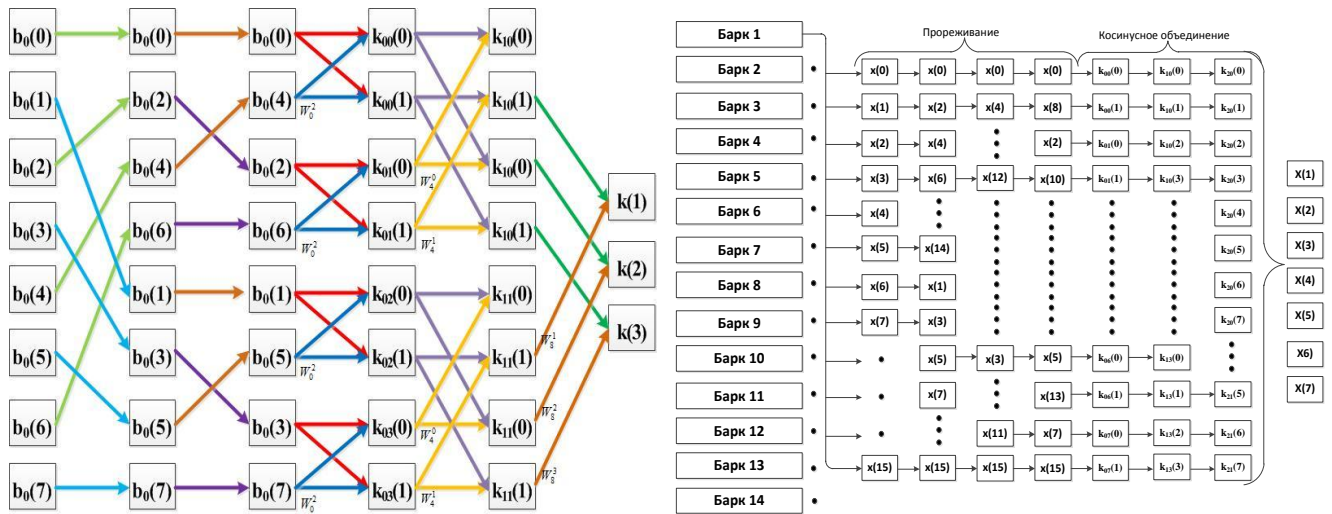


Рисунок 1 – Граф адаптированных алгоритмов БПФ (слева) и ДКП (справа)

$$b(k) = h(n) \cdot \sqrt{\frac{1}{N}} \cdot \sum_{n=1}^{2N-1} x(n) \cdot \cos\left(\frac{2\pi(k+1) \cdot (n+1)}{4N}\right) \quad (2)$$

$$h(n) = \sin^4\left[\frac{\pi \cdot \sin(\pi/2) \cdot (2n+1)}{4N}\right] \quad (3)$$

Зеркальное преобразование входного сигнала относительно нулевого отсчёта (4) позволяет достичь минимальных искажений при выполнении половинного перекрытия окон (рисунок 2).

$$x'(j) = \begin{cases} x(N-n-1), & \text{если } j < N \\ x(n), & \text{в остальных случаях} \end{cases} \quad (4)$$

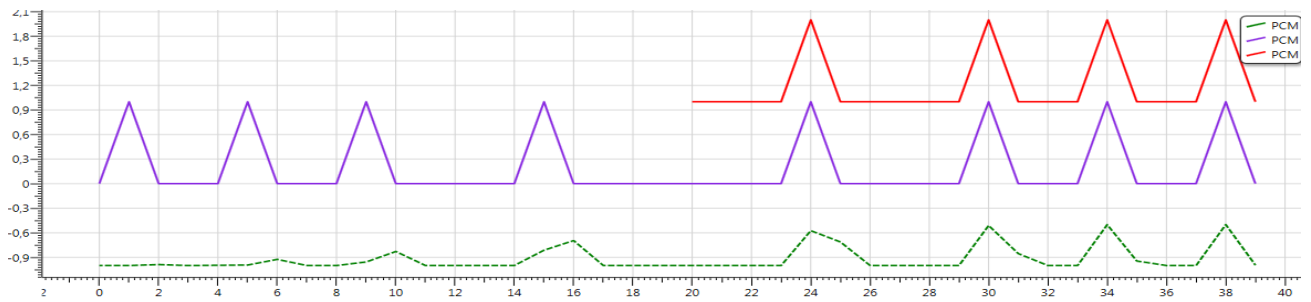


Рисунок 2 – график входного (-), зеркального (-) и выходного (--) сигнала косинусного преобразования с применением синусного окна

При уменьшении масштаба частотных отсчётов пропорционально растёт степень сжатия и незначительно ухудшается качество воспроизведения. Пороговым значением масштабирования для разборчивости является единичное значение частоты. В результате выполнения ортогональных преобразований достигается степень сжатия до четырёх раз.

Адаптированные для звука методики сжатия без потерь информации увеличат коэффициент компрессии при неизменности качества, однако внесут задержку в процедуру передачи данных, что может повлиять на комфортность общения. Производными от оригинальных параметров и объектов сжатия выступают значения частот и амплитуд каждого тона отдельного барка. Таким образом, входными данными для проведения компрессии являются кванты частоты. Описание алфавитов символов (литералов) и длин (смещений) образуют структуры лексикографически построенных кодовых деревьев (см. таблицу 1).

Таблица 1 - Определение диапазонов и типов данных кодов для сжатия без потерь

Алфавит длин	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Длина кода символа	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC
Код в дереве	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN
№ литерала из алфавита частоты (барки)						Размер кода из дерева размера					Код в дереве								
0 350 Гц						DEC					{BIN}								
1 450 Гц						DEC					{BIN}								
2 570 Гц						DEC					{BIN}								
3 700 Гц						DEC					{BIN}								
...													
256 4000 Гц						DEC					{BIN}								
Символ алфавита амплитуд	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14				
Длина кода символа	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC				
Выходной код	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN				
Символ алфавита амплитуд	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29				
Длина кода символа	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC	DEC				
Выходной код	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN	BIN				

Здесь номера литералов – это параметры частотного сигнала, длины кодов DEC которых определены путём построения двухуровневых кодовых деревьев, содержащих сжатые значения BIN (рисунок 3).

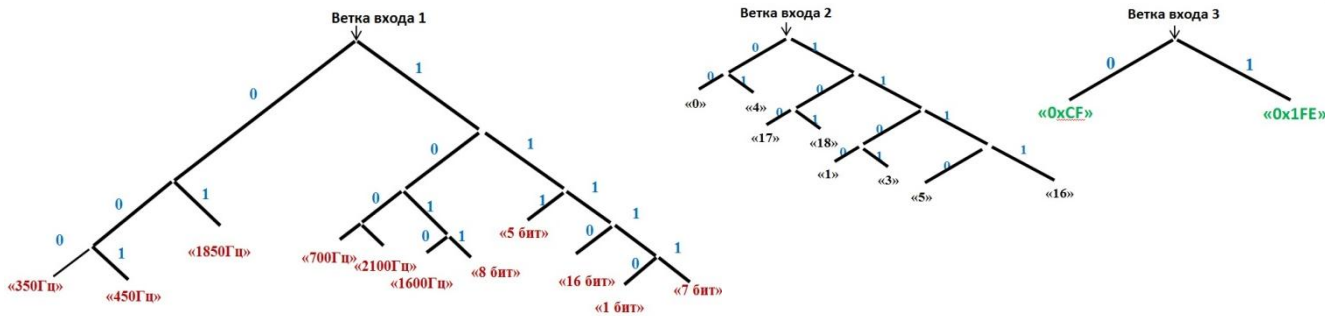


Рисунок 3 – Деревья для кодов амплитуд, длин смещений и повторов

Процесс компрессии в совокупности с процедурами линейного и ортогонального преобразования позволяет сократить объем данных в 14 раз.

В третьей главе осуществлён синтез математической модели для функциональной и алгоритмической реализации устройства динамического кодера. В области линейной обработки построен псевдо-квадратурный зеркальный банк фильтров, который разделяет широкополосный сигнал на критичные диапазоны с минимальными фазовыми сдвигами. Импульсная характеристика фильтра (5) представляет вектор свёртки для входного сигнала.

$$h_n = \begin{cases} \frac{\psi \times p_n \times (\sin(\omega_2 \times n) - \sin(\omega_1 \times n))}{\pi \times n}, & \text{если } n \geq 1 \\ \omega_0, & \text{в остальных случаях} \end{cases}$$

$$p_n = 1 - \frac{n}{\theta \times N} \quad (5)$$

Проверка каждого фильтра проводилась по методике вычисления минимальной дисперсии банка фильтров, используя неунифицированное ДПФ. Если величина дисперсии остается в пределах входных параметров математического модели психоакустического анализа, то можно сделать вывод о приемлемости использования синтезированной структуры.

На графиках рисунка 4 представлен результат линейного преобразования входного униполярного сигнала в свёрнутый биполярный сигнал, что является неотъемлемой частью подготовки звуковой информации ЭВМ к дискретным ортогональным преобразованиям.

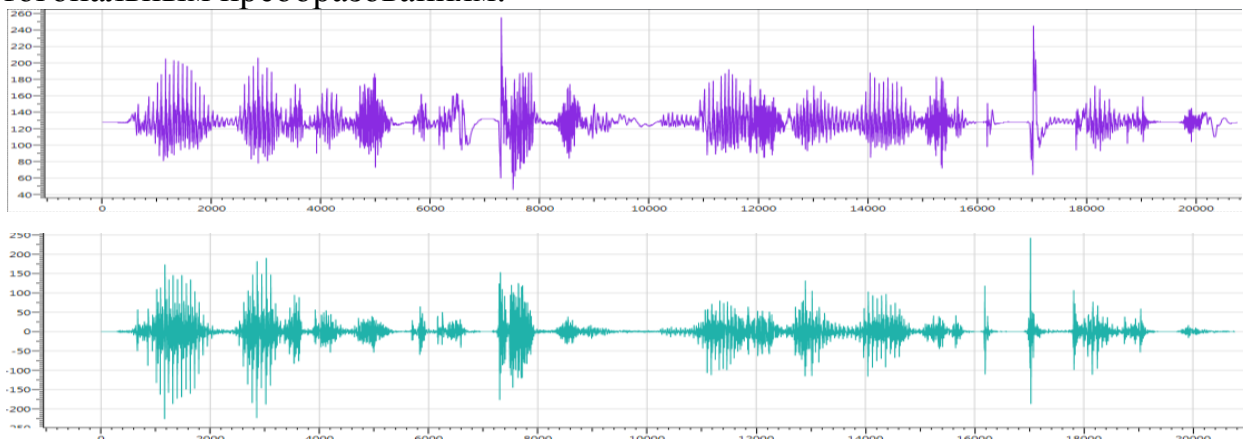


Рисунок 5 – Исходный речевой сигнал (вверху) и преобразованный (внизу)

Ортогональные преобразования временного речевого потока реализуют спектр тональностей в широком и узком диапазонах и осуществляют организацию спектральных массивов для получения и наложения психоакустической маски. Вариация коэффициентами математических функций восприятия звука (6,7,8) задаёт его амплитудно-частотные характеристики, которые содержат основные форманты речи в текущем временном отрезке, определяющие выразительность и уникальность голоса (рисунок 5).

$$T(f) = \alpha \times \left(\frac{f}{1000}\right)^{-0,8} - \beta \times e^{-0,6 \times \left(\frac{f}{1000} - 3,3\right)^2} + \gamma \times \left(\frac{f}{1000}\right)^4 \quad (6)$$

$$SF(x) = \varepsilon + \Omega \times (x + \mu) - \Theta \times \sqrt{1 + (x + \mu)^2} \quad (7)$$

$$W(t) = (1 - \omega) \left(1 + \frac{2t}{\varphi}\right) \exp\left(-\frac{2t}{\varphi}\right) + \omega \left(1 + \frac{2t}{\sigma}\right) \exp\left(-\frac{2t}{\sigma}\right) \quad (8)$$

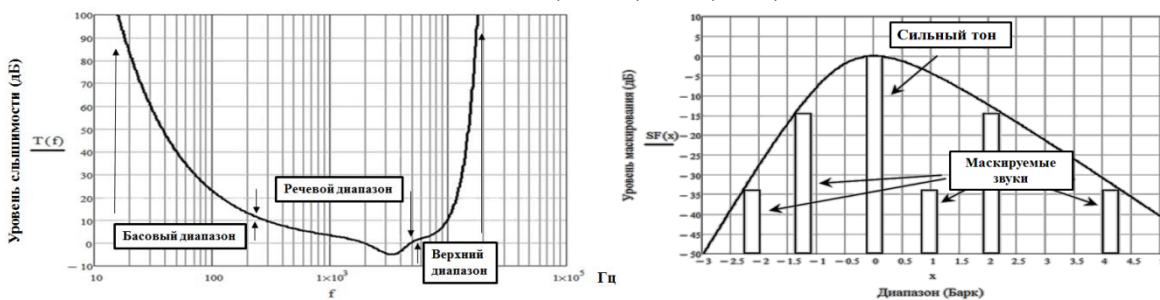


Рисунок 5 – Маскирование восприятия речевого сигнала

Алгоритмы наложения частотно-временных масок (рисунок 6) выполняют основную роль в области сжатия с потерями, от вариации их параметров зависят как степень компрессии, так и качество воспроизводимого звука.

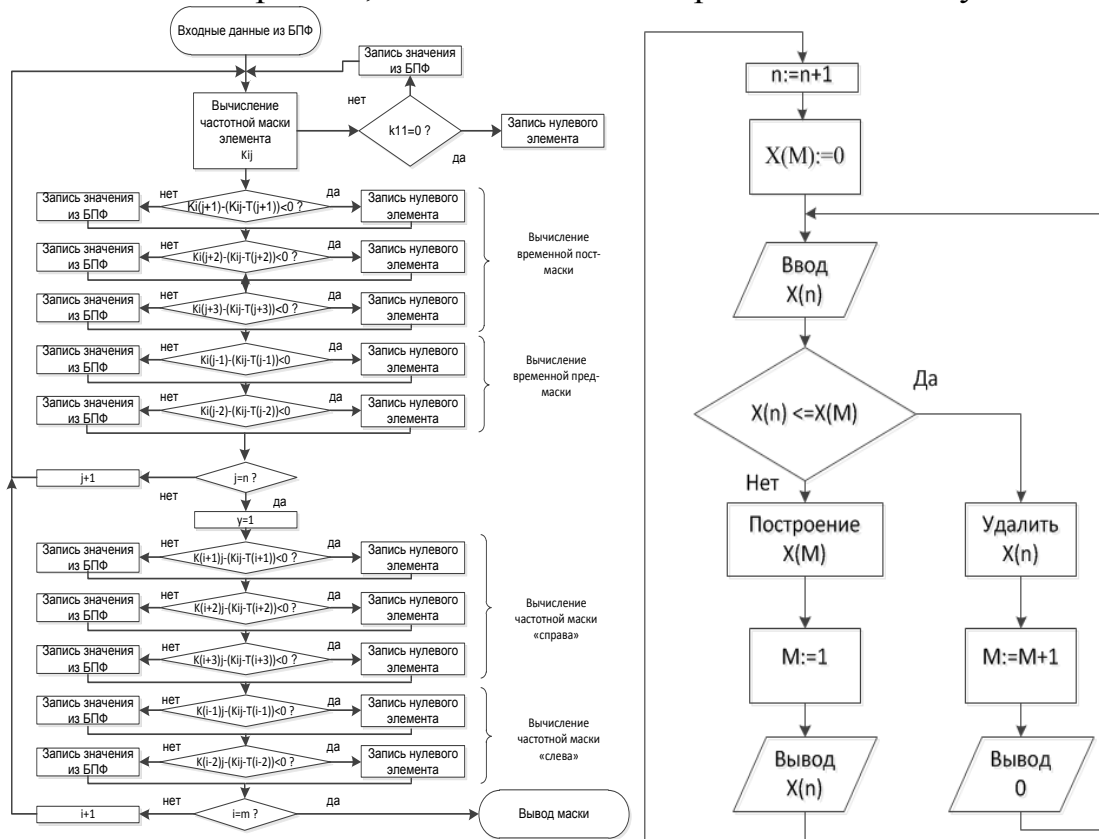


Рисунок 6 – Алгоритмы частотной (слева) и временной (справа) масок

Нахождение маски и результат прореживания по алгоритму наложения частотной маски представлены на графиках рисунка 7.

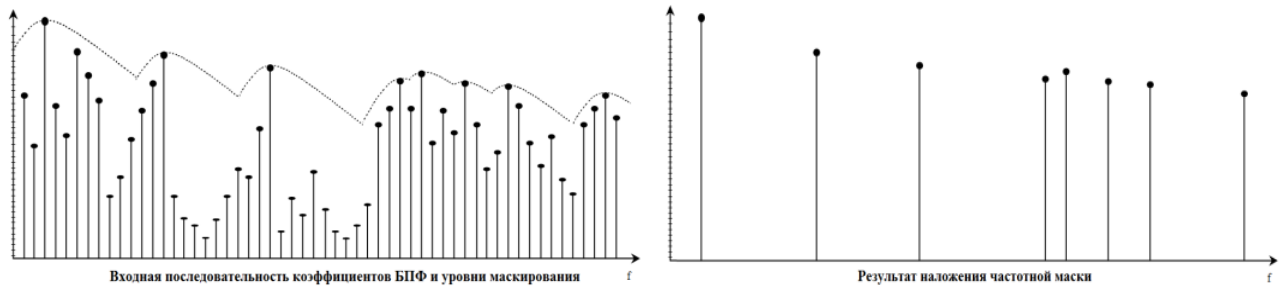


Рисунок 9 - Нахождение частотной маски (слева) и наложение (справа)

Выявленные особенности произношения в русском языке создали основу для повышения плотности информативных составляющих при передаче речевого потока путём замены редуцированных звуков шумовым формантным сигналом диктора.

Построенная модель наделяет средство сжатия параметрическими и функциональными взаимосвязями, что позволяет говорить о её динамичности и, при изменении методической и алгоритмических операций, – о её вариативности.

В четвертой главе выполнен структурный синтез адаптивного кодера и средства оценки его качественных свойств. Наличие результатов исследований математических моделей и набора методик сжатия позволило осуществить синтез структуры кодера в целом, а также определить в ней вариативные параметры и степень их влияния на изменение речевого потока, установить степень динамичности цепей внутри структуры.

С целью объединения перцептивных методов обработки звукового потока он был разделён на параллельные ортогональные преобразования (Рисунок 10).

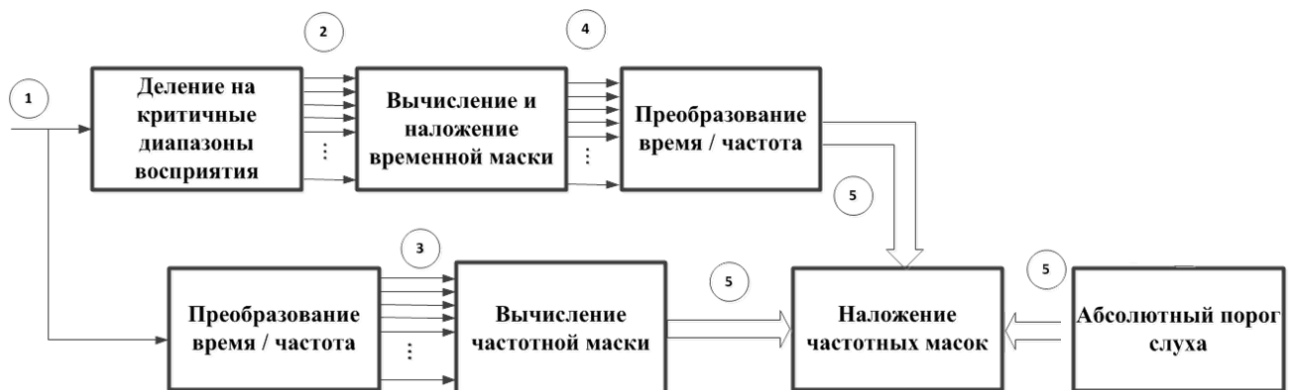


Рисунок 10 – Параллельные ортогональные процессы

В первом осуществлялись процессы подготовки к вычислению и наложению временной маски, во втором – подготовка и вычисление частотной маски. В результате выполняется взаимодействие трёх моделей: частотной, временной и абсолютной. Поток №1 представляет собой вектор дискретных отсчётов аналого-цифрового преобразования вида $b = [b_0, b_1, b_2, b_3, \dots, b_n]$ с разрядностью n , определённой

точкой изгиба q (9) передаточной характеристики неравномерного квантования с вычисленной разрядностью m при заданном динамическом диапазоне D .

$$b_i = \begin{cases} \frac{A \cdot x_i \cdot 2^n}{2^m}, & \text{для } x \leq \frac{1}{A} \\ \frac{1 + \ln(A \cdot \frac{x_i}{2^m})}{1 + \ln A} \cdot 255, & \text{для } x > \frac{1}{A} \end{cases}; \quad A = 2^m \div q; \quad m = \log_2 10^{D/10} \quad (9)$$

Поток №2 содержит параллельные векторы \hat{b} (матрицу 10), отражающие критичные полосы. Преобразование «время/частота» происходит по адаптированному алгоритму ДКП. Результатом спектрального преобразования №4 является матрица \hat{f} , количество столбцов которой уменьшено вдвое по сравнению с \hat{b} . Выходные данные №3 представляют собой массивы спектров для анализа окон маскирования по частоте.

$$\hat{b} = \begin{bmatrix} \hat{b}_{00} & \hat{b}_{01} & \hat{b}_{02} & \dots & \hat{b}_{0n} \\ \hat{b}_{10} & \hat{b}_{11} & \hat{b}_{12} & \dots & \hat{b}_{1n} \\ \hat{b}_{20} & \hat{b}_{21} & \hat{b}_{22} & \dots & \hat{b}_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{b}_{m0} & \hat{b}_{m1} & \hat{b}_{m2} & \dots & \hat{b}_{mn} \end{bmatrix} \quad \hat{f} = \begin{bmatrix} \hat{f}_{00} & \hat{f}_{01} & \hat{f}_{02} & \dots & \hat{f}_{0(n/2)} \\ \hat{f}_{10} & \hat{f}_{11} & \hat{f}_{12} & \dots & \hat{f}_{1(n/2)} \\ \hat{f}_{20} & \hat{f}_{21} & \hat{f}_{22} & \dots & \hat{f}_{2(n/2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{f}_{m0} & \hat{f}_{m1} & \hat{f}_{m2} & \dots & \hat{f}_{m(n/2)} \end{bmatrix} \quad (10)$$

Результатом наложения маски (потоки №5) является спектр (11), который содержит основные форманты речи в текущем временном отрезке, определяющие выразительность и уникальность воспроизведения голоса.

$$b = [b_0, b_1, b_2, b_3, \dots, b_n] \Rightarrow \begin{bmatrix} k_0 \\ k_1 \\ k_2 \\ \vdots \\ k_{n/2} \end{bmatrix} \Rightarrow \begin{bmatrix} k_0 \cdot z^{-m} \\ k_1 \cdot z^{-m} \\ k_2 \cdot z^{-m} \\ \vdots \\ k_{n/2} \cdot z^{-m} \end{bmatrix} + \dots + \begin{bmatrix} k_0 \cdot z^{-1} \\ k_1 \cdot z^{-1} \\ k_2 \cdot z^{-1} \\ \vdots \\ k_{n/2} \cdot z^{-1} \end{bmatrix} + \begin{bmatrix} k_0 \\ k_1 \\ k_2 \\ \vdots \\ k_{n/2} \end{bmatrix} \Rightarrow M = \begin{bmatrix} k_{11} & k_{12} & \dots & k_{1(n-1)} & k_{1n} \\ k_{21} & k_{22} & \dots & k_{2(n-1)} & k_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ k_{(m-1)1} & k_{(m-1)2} & \dots & k_{(m-1)(n-1)} & k_{(m-1)n} \\ k_{m1} & k_{m2} & \dots & k_{m(n-1)} & k_{mn} \end{bmatrix} \quad (11)$$

Итогом структурного проектирования кодера является схема его функционального взаимодействия, представленная на рисунке 11.

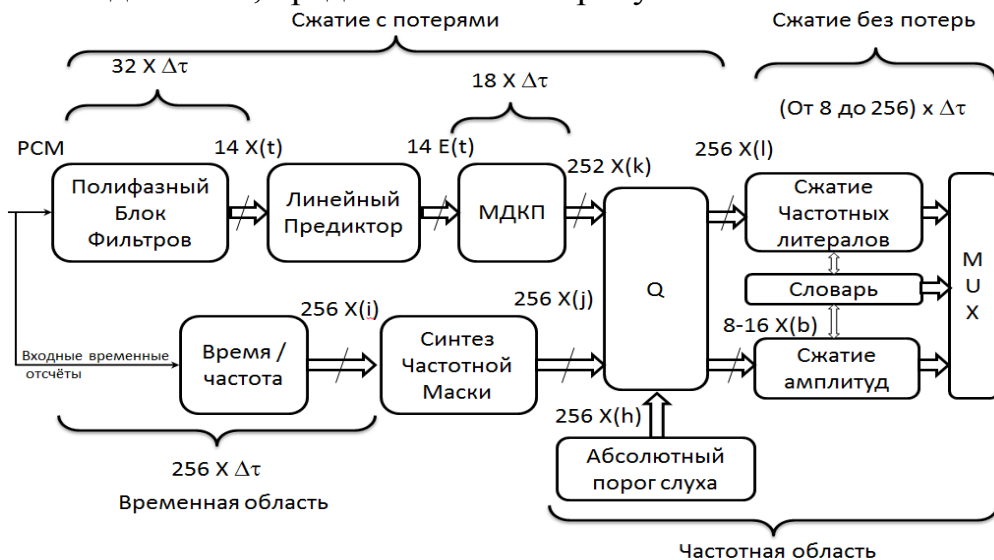


Рисунок 11 – Структура синтезированного кодера

Определение временной сложности цифровых преобразований заключается в анализе времени преобразований ортогональных блоков и цифровой фильтрации. В синтезированной функциональной модели поток входных оцифрованных данных попадает параллельно на модифицированный полифазный фильтр, имеющий динамическую структуру, в нашем случае – 14 полосовых фильтров, и на блок адаптивного БПФ от 256 отсчётов. В блоке «линейного предиктора» происходит наложение временной маски. Модифицированное дискретное косинусное преобразование производит ортогональное преобразование «время-частота». Квантователь обеспечивает масштабирование трёх спектров: модели абсолютного порога слуха, психоакустической маски, вычисленной из АБПФ и входных частотных значений. Далее параллельный поток подвергается сжатию без потерь.

Вариативными блоками в кодере определены: банк полифазных фильтров (с динамикой выбора сворачивающего вектора и количества критичных полос); линейный предиктор (с различной длительностью выборки и функцией временного маскирования); функционал квантователя (с изменяемой глубиной частотного маскирования); блок сжатия без потерь (с выбором техники сжатия между статическим и динамическим словарём, а также методом скользящего окна).

Эффект от применения процедур сжатия показан в таблице 2.

Таблица 2 - Длина маски, разрядность АЦП и эффективности сжатия.

Разрядность шкалы мощности	Динамический диапазон, дБ	Максимальная степень сжатия	Компрессия, бит удалено (Окно)		
			8КГц	22,05КГц	44,1КГц
8 бит	24	14 (7%)	320	882	1764
12 бит	36	20 (5%)	480	1323	2646
16 бит	48	35 (<3%)	800	2205	4410
24 бит	72	>50 (2%)	1280	3308	6615

Сжатие путём наложения частотной маски позволило осуществить компрессию с соотношением до 14:1 для речевого сигнала, что подтверждено работоспособностью программной имплементации кодера на ЭВМ.

Соответствие процентного отношения сжатых данных различных техник сжатия без потерь с Deflate при разной скорости показано в таблице 3.

Таблица 3 - Методики сжатия без потерь при разной скорости.

Частота дискретизации, Гц	Код, бит	Эффективность сжатия, %		
		Статический словарь	Динамический словарь	Скользящее окно
8 000	8	99%	127%	96 %
	12	99%	110%	94%
	16	98%	107%	90%
	24	97%	106%	84%
22 500	8	92%	106%	95%
	12	90%	101%	92%
	16	89%	95%	91%
	24	88%	89%	90%
44 100	8	90%	89%	95%
	12	90%	87%	92%
	16	90%	85%	91%
	24	90%	85%	90%

Спектральные искажения при работе кодека определены сопоставлением БПФ для семплов исходного и переданного потоков. В случае, если спектральная плотность мощности шума не превышает граничное значение, заданное проектировщиком, оговаривается приемлемость разборчивости речевого сигнала.

Для проверки спроектированных структур и методик сжатия речевого потока реализовано программное обеспечение, которое позволяет произвести анализ выходного звука на разборчивость и выразительность (рисунок 12).

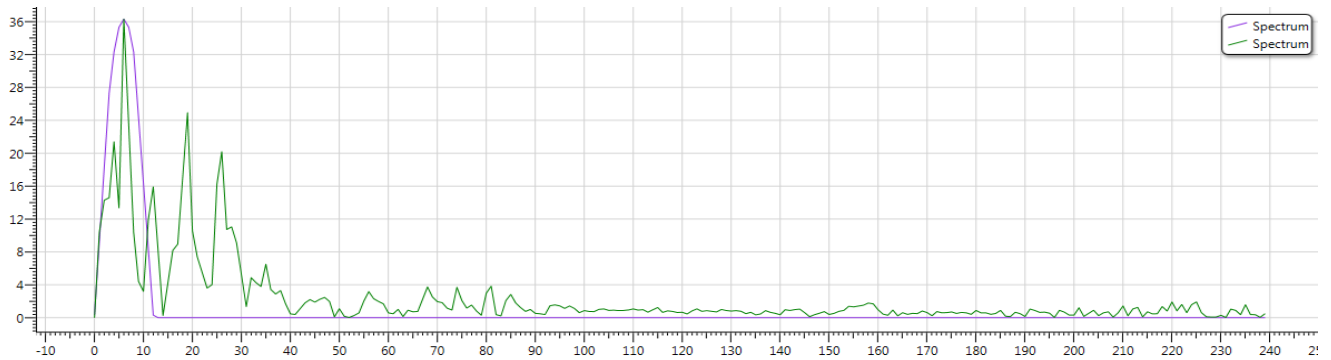


Рисунок 12 – Спектр анализируемого (-) сигнала и наложенная на него маска (-)

Имплементация предоставляет возможность провести обратное ортогональное преобразование и оценить визуально результат на графике, содержащем временное представление сигнала, а также прослушать оригинальную звуковую дорожку и результат.

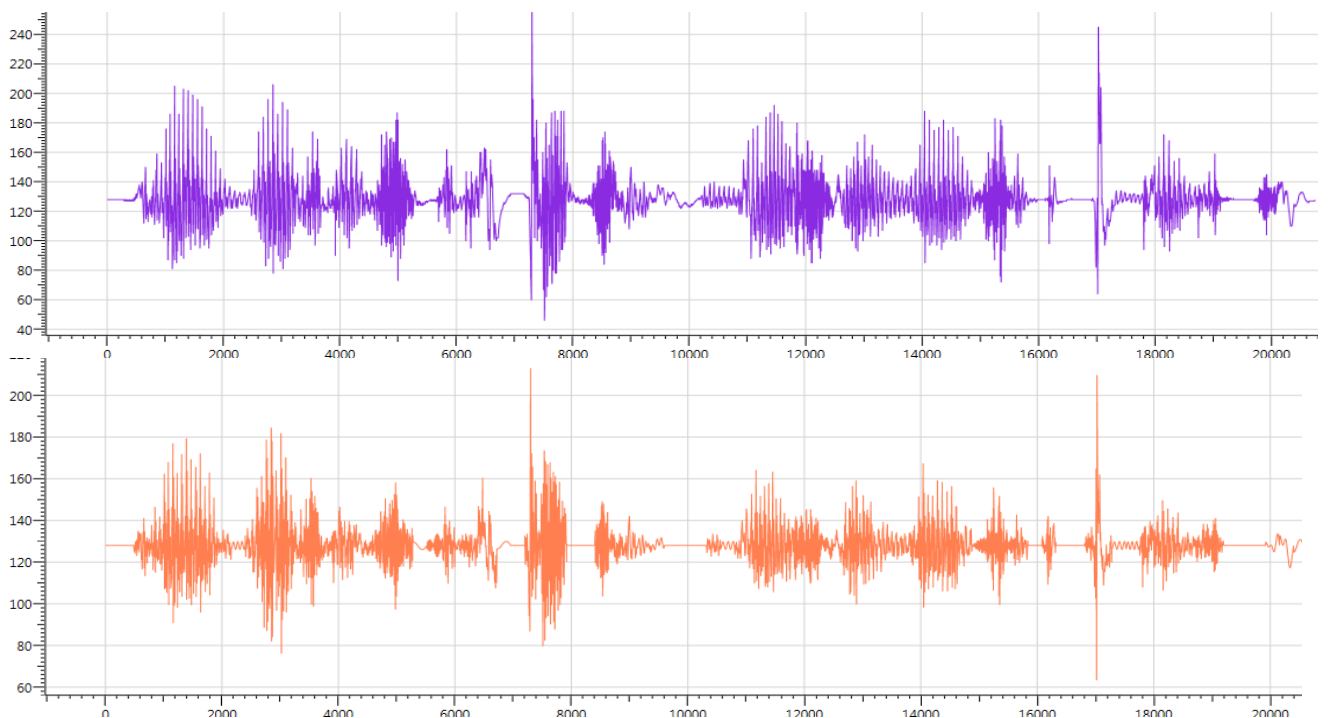


Рисунок 14 – Сравнение сигнала до сжатия (-) и сигнала после декодирования (-) при степени сжатия 14:1

Также был проведён эксперимент по определению разборчивости среди экспертных слушателей, который показал, что порог разборчивости речи находится значительно ниже порога узнаваемости личности диктора, и составляет аппроксимировано 325 Байт/с в первом случае и 520 Байт/с во втором, что ниже значений современных речевых кодеров G.729, CELT, ADPCM.

Результаты экспериментального исследования показали, что вариативность, т.е. выбор процедур и методик сжатия, для построения речевого кодера зависит от приоритета поставленной задачи в системе хранения и передачи: разборчивости текста или узнаваемости (выразительности) говорящего.

ЗАКЛЮЧЕНИЕ

Основным результатом работы является многофункциональный инструмент цифровой обработки речевого сигнала с изменяемым уровнем сжатия и битрейтом, а также с функциями оценки задержки, который является базой для построения семейства кодеров с параметрами, необходимыми для внедрения в системы хранения и дуплексного обмена речевой информацией реального времени. По результатам исследований можно сделать следующие выводы:

- анализ техник речевого кодирования показал основные преимущества ортогонального и линейного метода компрессии, в результате было выбрано гибридное решение для синтеза структуры кодера. Процедуры были разделены на две части: блок сжатия с потерями и без. Блок сжатия с потерями, в свою очередь, разделён на линейные и ортогональные преобразования;

- в процессе линейной цифровой обработки были синтезированы математические модели блока полифазной фильтрации и линейного предиктора, построены алгоритмы их реализации. Выявлено, что данные блоки позволяют управлять критериями разборчивости и выразительности речи;

- проведена разработка модели и алгоритма наложения психоакустической маски – основного ортогонального блока, реализующего сжатие с потерями. Данный блок позволил варьировать битрейт на основе метода удаления неслышимых частотных компонентов, сжатие – до 4 раз;

- структурирована система сжатия без потерь по словарным и оконным методам, основной задачей которой является компрессия частотных выборок. Данная процедура в совокупности с предыдущим преобразованием позволила сжать поток до 14 раз;

- реализована программная имплементация речевого кодера, позволяющая применить методы и алгоритмы диссертационного исследования к записанному голосовому потоку, вычислить степень сжатия и задержку и оценить качество воспроизведения речи.

Полученные результаты соответствуют п. 4 «Разработка методов и алгоритмов решения задач системного анализа, оптимизации, управления, принятия решений и обработки информации», п. 7 «Методы и алгоритмы структурно-параметрического синтеза и идентификации сложных систем» паспорта специальности 05.13.01 «Системный анализ, управление и обработка информации (технические системы)».

В развитие итогов исследования перспективными задачами представляются: нахождение составных базовых функций ортогональных преобразований, содержащих универсальные энергетические сочетания голоса диктора; исследование путей автоматического выбора процедур сжатия и их параметров при изменении условий передачи сигнала; создание голосовых лексикографических словарей, увеличивающих плотность сжатия.

СПИСОК РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ

В изданиях, рекомендованных ВАК Минобрнауки РФ:

1. Шиленков, Е.А. Способ построения частотно-временной маски для речевого кодирования на основе психоакустической модели слуха / Е.А. Шиленков // Проектирование и технология электронных средств. – 2013. - №4. Владимир – С. 16-20.
2. Потапенко, А.М., Шиленков, Е.А. Методика построения динамических сигнально-кодированных конструкций с целью повышения пропускной способности канала связи / А.М. Потапенко, Е.А. Шиленков // Проектирование и технология электронных средств. – 2012. - №1. Владимир – С. 12-16.
3. Шиленков, Е.А. Нетривиальность анализа сигналов в дискретном канале / Е.А. Шиленков, К.Ю. Рюмшин, А.Ю. Богомазов // Известия ЮЗГУ. – 2012. - №2 (Ч.2). Курск – С. 72-76.
4. Шиленков, Е.А., Сравнение различных методов построения недвоичных низкоплотностных кодов / Е.А. Шиленков, К.А. Воробьев // Известия ЮЗГУ. – 2012. - №2 (Ч.3). Курск – С. 138-144.
5. Цыплаков, Ю.В., Шиленков, Е.А. Сравнительный анализ специальных средств для измерения реального затухания электромагнитных волн и контроля качества средств экранирующей групповой защиты / Е.А. Шиленков, Ю.В. Цыплаков // Известия ЮЗГУ. – 2012. - №2 (Ч.3). Курск – С. 32-37.
6. Воробьев, К.А., Шиленков, Е.А. Широкополосные многочастотные сигнально-кодированные конструкции для передачи информации через ионосферный канал / К.А. Воробьев, И.С. Косилов, Е.М. Лобов, Е.А. Шиленков // Известия ЮЗГУ. – 2012. - №2 (Ч.3). Курск – С. 41-46.
7. Мухин, И.Е., Шиленков, Е.А. Метод параметрического синтеза антенно-фидерных, радиоприёмных и демодуляторных средств сигнально-приёмного тракта современных систем телекоммуникаций / И.Е. Мухин, Е.А. Шиленков // Известия ЮЗГУ. – 2012. - №2 (Ч.3). Курск – С. 110-115.

В других изданиях:

8. Шиленков, Е.А. Кодирование и сжатие речевого сигнала в ТФОП / Е.А. Шиленков, А.М. Потапенко // Общество, современная наука и образование: проблемы и перспективы: сб. науч. тр. по мат-лам Междунар. заоч. науч.-практ. конф. - 2012: (Ч.10). Тамбов, - С. 159-161.
9. Шиленков, Е.А., Хотынюк, С.С. Линейный предиктор в ортогональном речевом кодере / Е.А. Шиленков, С.С. Хотынюк // Материалы Всероссийской научно-практической конференции «Инфокоммуникации и информационная безопас-

- ность: состояние, проблемы и пути решения», Курск, 25-26 апреля 2014. – С. 11 - 17.
10. Шиленков, Е.А. Методика дескриптора LZSTACKER / Е.А. Шиленков // Материалы Всероссийской научно-практической конференции «Инфокоммуникации и информационная безопасность: состояние, проблемы и пути решения», Курск, 25-26 апреля 2014. – С. 18 - 22.
11. Шиленков, Е.А. Методика поиска сжатых блоков в поточных данных / Е.А. Шиленков // Материалы Всероссийской научно-практической конференции «Инфокоммуникации и информационная безопасность: состояние, проблемы и пути решения», Курск, 25-26 апреля 2014. – С. 22 - 24.
12. Ефимова, О.В., Шиленков, Е.А. Банк фильтров ортогонального речевого кодирования / О.В. Ефимова, Е.А. Шиленков // Материалы Всероссийской научно-практической конференции «Инфо-коммуникации и информационная безопасность: состояние, проблемы и пути решения», Курск, 25-26 апреля 2014. – С. 113 - 118.
13. Шиленков, Е.А. Методика дескриптора данных по динамическому словарю / Е.А. Шиленков // Материалы Всероссийской научно-практической конференции «Инфокоммуникации и информационная безопасность: состояние, проблемы и пути решения», Курск, 25-26 апреля 2014. – С. 118 - 122.
14. Белугин, И.Н., Шиленков, Е.А. Повышение структурной устойчивости речевого ортогонального кодера / И.Н. Белугин, Е.А. Шиленков // Материалы Всероссийской научно-практической конференции «Инфокоммуникации и информационная безопасность: состояние, проблемы и пути решения», Курск, 25-26 апреля 2014. – С. 232 - 236.
15. Шиленков, Е.А. Методика дескриптора данных по статическому словарю / Е.А. Шиленков // Материалы Всероссийской научно-практической конференции «Инфокоммуникации и информационная безопасность: состояние, проблемы и пути решения», Курск, 25-26 апреля 2014. – С. 285 – 287.
16. Шиленков, Е.А. Определение дескриптора и формата словаря сжатых данных / Е.А. Шиленков // Материалы Всероссийской научно-практической конференции «Инфокоммуникации и информационная безопасность: состояние, проблемы и пути решения», Курск, 25-26 апреля 2014. – С. 287-289.
17. Матвеевко, В.В., Шиленков, Е.А. Анализ способов создания сигнальных конструкций для систем радиосвязи повышенной энергетической скрытности / В.В. Матвеевко, Е.А. Шиленков // Актуальные проблемы инфотелекоммуникаций: Сборник материалов региональной конференции / Юго-Зап. гос. ун-т. Курск, 2011., стр.11
18. Матвеевко, В.В., Шиленков, Е.А. Способ приема сигналов с внутрисимвольной псевдослучайной перестройкой рабочей частоты / В.В. Матвеевко, Е.А. Шиленков // Актуальные проблемы инфотелекоммуникаций: Сборник материалов региональной конференции / Юго-Зап. гос. ун-т. Курск, 2011., стр.7
19. Матвеевко, В.В., Бабанин, И.Г., Шиленков, Е.А. Способ построения быстродействующих цифровых фильтров для защиты широкополосных радиоприемных устройств / В.В. Матвеевко, И.Г. Бабанин, Е.А. Шиленков // Актуальные

проблемы инфотелекоммуникаций: Сборник материалов региональной конференции / Юго-Зап. гос. ун-т. Курск, 2011.,стр.98

20. Матвеевко, В.В., Шиленков, Е.А. Передатчик для сигнально-кодовых конструкций с расширением спектра методом псевдо-случайной перестройки рабочей частоты / В.В. Матвеевко, Е.А. Шиленков // Актуальные проблемы инфотелекоммуникаций: Сборник материалов региональной конференции / Юго-Зап. гос. ун-т. Курск, 2011.,стр.100

Подписано к печати __.__.2016

Объём 20 печ. л. Тираж 80 экз.

Отпечатано в _____