

На правах рукописи



**Болдышев Алексей Владимирович**

**РАЗРАБОТКА И ИССЛЕДОВАНИЕ МЕТОДОВ И АЛГОРИТМОВ  
СУБПОЛОСНОГО КОДИРОВАНИЯ РЕЧЕВЫХ СООБЩЕНИЙ ПРИ  
ХРАНЕНИИ И ПЕРЕДАЧЕ РЕЧЕВЫХ ДАННЫХ**

по специальности 05.13.17 – Теоретические основы информатики

**АВТОРЕФЕРАТ**  
диссертации на соискание ученой степени  
кандидата технических наук

Белгород – 2013

Работа выполнена в Федеральном государственном автономном образовательном учреждении высшего профессионального образования «Белгородский государственный национальный исследовательский университет» (НИУ «БелГУ»), на кафедре информационно-телекоммуникационных систем и технологий

Научный руководитель **Жиляков Евгений Георгиевич**, доктор технических наук, профессор

Официальные оппоненты: **Нечаев Юрий Борисович**, доктор физико-математических наук, профессор, заслуженный деятель науки РФ, Воронежский государственный университет, профессор кафедры информационных систем, г. Воронеж

**Белов Александр Сергеевич**, кандидат технических наук, ФГАОУ ВПО «Белгородский государственный национальный исследовательский университет» доцент кафедры прикладной информатики, г. Белгород

Ведущая организация Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Белгородский государственный технологический университет им. В.Г. Шухова»

Защита состоится «3» апреля 2013 года в 16 часов 30 минут на заседании диссертационного совета Д 212.015.10 на базе ФГАОУ ВПО «Белгородский государственный национальный исследовательский университет» по адресу: 308015 г. Белгород, ул. Победы, д. 85, ауд. 3-8, к. 15

С диссертацией можно ознакомиться в научной библиотеке ФГАОУ ВПО «Белгородский государственный национальный исследовательский университет» (НИУ «БелГУ») по адресу: 308015 г. Белгород, ул. Победы, д. 85.

Автореферат разослан «\_\_\_» февраля 2013 г.

Ученый секретарь  
диссертационного совета, к.т.н., с.н.с.



Белов С.П.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность диссертационного исследования.** Информационный обмен является важной и неотъемлемой частью человеческой деятельности (социальной, экономической), о чем свидетельствует постоянное развитие информационно-телекоммуникационных технологий обработки, хранения и передачи данных. В настоящее время происходит повышение интенсивности информационного обмена на основе речевых сообщений. Это обусловлено тем, что речевые сообщения являются наиболее естественной и удобной формой информационного обмена.

В информационно-телекоммуникационных системах (ИТС) передача и хранение речевых сообщений осуществляется в виде специальных кодовых комбинаций, совокупность которых естественно называть речевыми данными.

Одной из основных проблем реализации информационного обмена является ограниченность ресурсов ИТС (пропускная способность канала связи для передачи данных, объемы памяти информационных хранилищ). Поэтому, не вызывает сомнения, необходимость выбора такого способа кодирования, который обеспечивает минимум объемов битовых представлений хранимых и передаваемых речевых данных при сохранении приемлемого, с точки зрения пользователя, качества воспроизведения исходных речевых сообщений.

Решением проблемы минимизации объемов битовых представлений речевых данных занималось большое количество ученых, среди которых следует отметить: Б. Голда, Е.Г. Жиликова, Э. Оппенгейма, А.А. Пирогова, Л. Р. Рабинера, В.Г. Санникова, М.А. Сапожкова, Р.В. Шафера, О.И. Шелухина и других российских и зарубежных ученых.

На сегодняшний день известно большое количество методов кодирования речевых сообщений, которые используются в системах IP-телефонии, а также при передаче речи по сетям сотовой связи (G.729, G.723.1, G.728, AMR, FR, и т.д.). Разработчики этих кодеков достигли определенного предела в показателях эффективности кодирования (сжатие до 10 раз) при сохранении приемлемого для пользователя качества воспроизведения.

Дальнейший прогресс повышения эффективности кодирования может быть достигнут на основе оптимального учета свойства сосредоточенности энергии речевых сигналов в полосе частот, составляющих малую долю от частоты дискретизации. Методы кодирования речевых сообщений с позиции разбиения общей полосы частот на отдельные частотные интервалы представляется естественным называть субполосными.

Из известных методов субполосного кодирования можно отметить форматы MP3, OGG и т.д., в основе которых используется дискретное преобразование Фурье (ДПФ), в том числе дискретное косинус

преобразование Фурье (ДКПФ). Однако, они приспособлены в основном для обработки музыкальных данных, что не позволяет учитывать наличие в речевых сообщениях пауз, кроме того, они не отвечают никаким критериям оптимальности, например, минимизации ширины адаптивно определяемой полосы частот, в которой сосредоточена основная доля энергии речевого сигнала, погрешности аппроксимации в ней трансформанты Фурье исходного сигнала и погрешности квантования по уровню коэффициентов этой аппроксимации.

Поэтому разработка оптимальных субполосных методов кодирования речевой информации, позволяющих минимизировать объемы битовых представлений речевых данных при условии сохранения заданного уровня информативности восстанавливаемых речевых сообщений, является актуальной задачей.

**Целью работы является** совершенствование методов и алгоритмов субполосного кодирования речевой информации с точки зрения минимизации объемов битовых представлений речевых данных при условии сохранения заданного уровня информативности восстанавливаемых речевых сообщений.

Для достижения цели были сформулированы и решены следующие **задачи**:

1. Анализ основных направлений развития методов кодирования речевых данных при хранении и передаче речевых сообщений.
2. Разработка и исследование метода и алгоритмов многополосного кодирования и декодирования речевых сообщений, минимизирующих объемы битовых представлений речевых данных.
3. Разработка и исследование метода и алгоритмов кодирования и декодирования речевых сообщений с минимизацией объемов битовых представлений речевых данных на основе субполосной дискретизации речевых сигналов.
4. Разработка программных реализаций алгоритмов субполосного кодирования и восстановления речевых сообщений.

**Методы исследований** базируются на теории Фурье-анализа и синтеза сигналов, линейной алгебры, теории вероятностей и математической статистики, теории принятия статистических решений, вычислительных экспериментах.

**Научную новизну работы составляет следующее:**

1. Метод и алгоритм адаптивного выделения многополосных информационных компонент речевых сигналов, с минимальной суммарной шириной, которые позволяют уменьшить частотную избыточность речевых сигналов с точки зрения кодирования данных.
2. Метод и алгоритм многополосного кодирования и декодирования речевых сообщений на основе оптимальной аппроксимации трансформант

Фурье в адаптивно определяемой совокупности частотных интервалов, что соответствует требованию наилучшего воспроизведения речевых сообщений.

3. Метод и алгоритм кодирования и декодирования речевых сообщений на основе субполосной дискретизации огибающих информационных субполосных компонент речевых сигналов, позволяющий сократить объемы сохраняемых данных на основе уменьшения структурной избыточности.

**Практическая значимость работы определяется** алгоритмами субполосного кодирования речевой информации, которые позволяют минимизировать объемы битовых представлений речевых данных.

Полученные результаты используются в ООО «НПП «Сигнал» БелГУ», что подтверждается соответствующим актом, а также в учебном процессе подготовки бакалавров и магистров факультета КНиТ ФГАОУ ВПО «Белгородский государственный национальный исследовательский университет».

**Область исследования.** Содержание диссертации соответствует паспорту специальности 05.13.17 «Теоретические основы информатики» (технические науки) по следующим областям исследований:

п.3. Исследование методов и разработка средств кодирования информации в виде данных. Принципы создания языков описания данных, языков манипулирования данными, языков запросов. Разработка и исследование моделей данных и новых принципов их проектирования.

п.5. Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечения, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений.

**Связь с научными и инновационными программами.** Диссертационное исследование проводилось в рамках следующих программ фундаментальных, поисковых и инновационных исследований: ГК №14.740.11.0494 от 1 октября 2010г, ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы. Программа «У.М.Н.И.К.» Фонда содействия развитию малых форм предприятий в научно-технической сфере. Проект №10256. Программа «У.М.Н.И.К.» Фонда содействия развитию малых форм предприятий в научно-технической сфере. Проект №14228. АВИЦП «Развитие научного потенциала высшей школы (2009-2011 годы)» Проект № 2.1.2/9382.

**Положения, выносимые на защиту:**

1. Методы и алгоритмы субполосного кодирования и декодирования речевых сообщений, позволяющие минимизировать объемы битовых представлений речевых данных.

2. Структура пакета речевых данных, закодированных на основе разработанных процедур.

3. Рекомендации по использованию разработанных алгоритмов кодирования и декодирования речевых сообщений.

4. Результаты вычислительных экспериментов, иллюстрирующие обоснованность и достоверность выводов.

**Достоверность полученных результатов** и выводов обусловлена корректностью математических преобразований, отсутствием противоречий с известными теоретическими положениями и выводами, а также подтверждается результатами большого количества вычислительных экспериментов по обработке реальных речевых данных.

**Личный вклад соискателя.** Все изложенные в диссертационной работе результаты исследований получены либо соискателем лично, либо при его непосредственном участии.

**Апробация работы.** Результаты диссертационного исследования обсуждались на следующих научно-технических конференциях: Международная молодежная конференция «Прикладная математика, управление и информатика» (г. Белгород, 2012г); V Международная научно-техническая конференция «Информационные технологии в науке, образовании и производстве» (г. Орел, 2012г); 14-ая, 13-ая, 12-ая Международная конференция и выставка «Цифровая обработка сигналов и ее применение - DSPA» (г. Москва, 2012, 2011, 2010 гг.); 2-ая Международная научно-техническая конференция «Компьютерные науки и технологии» (г. Белгород, 2011г.); 14-ая Международная научно-техническая конференция «Медико-экологические информационные технологии – 2011» (г. Курск, 2011г.); XXIII - Международной научной конференции «Математические методы в технике и технологиях - ММТТ-23» (г. Саратов, 2010г.); 1-ая Международная научно-техническая конференция «Компьютерные науки и технологии» (г. Белгород, 2009г.).

**Публикации.** Основные положения диссертационной работы изложены в 20 печатной работе, из них 8 статей в журналах из списка ВАК. Получено 4 свидетельства Роспатента РФ об официальной регистрации программ для ЭВМ.

**Объем и структура работы.** Диссертация состоит из Введения, четырех глав, Заключение и Приложений. Работа изложена на 158 страницах машинописного текста, включая 30 рисунков, 37 таблиц и список литературных источников из 149 наименований.

## **СОДЕРЖАНИЕ РАБОТЫ**

**Во введении** обосновывается актуальность диссертационного исследования, формулируется основная цель и дается обзор содержания глав.

**Глава 1. Информационные процессы на основе речевых сообщений, тенденции, проблемы и методы реализации.** В главе основное внимание уделяется описанию областей и направлений использования информационно-телекоммуникационных систем (ИТС) для реализации

информационного обмена на основе речевых сообщений. Рассматривается проблема ограниченности ресурсов ИТС (пропускная способность каналов связи и объемы памяти информационных хранилищ), требуемых для передачи и хранения речевых данных. Рассматриваются основные известные подходы к кодированию речевых сообщений с целью минимизации объемов их битовых представлений при сохранении заданного уровня информативности.

Обработке подвергается речевой сигнал, под которым понимается результат реакции некоторых регистрирующих устройств на акустическое воздействие, генерируемое речевой системой человека. Предполагается, что они изначально дискретизируются с частотой  $f_d = 8\kappa\Gamma\text{ц}$ , разрядность АЦП равна 8.

Можно выделить два основных направления кодирования речевых данных с целью минимизации их объема, а именно кодирование сигнала в паузах (кодирование заключается в определении начала паузы и ее длительности) и кодирование данных, которые порождены присутствием звуков речи.

Дается обоснование целесообразности использования в создаваемых методах и алгоритмах кодирования речевых сообщений вариационных методов субполосного анализа, основанных на использовании субполосных матриц вида<sup>1</sup>:

$$A_r = \{a_{ik}^r\} = \int_{\omega \in W_r} e^{-j\omega(i-k)} d\omega = \frac{\sin[\Omega_{r+1}(i-k)] - \sin[\Omega_r(i-k)]}{\pi(i-k)}, \quad (1)$$

где  $W_r$  - некоторый интервал из исходной частотной оси

$$W_r = [-\Omega_{r+1}, -\Omega_r) \cup [\Omega_r, \Omega_{r+1}), \Omega_0 = 0, \quad (2)$$

где  $\Omega_r$  - нормированная круговая частота.

Отмечена целесообразность представления субполосных матриц (1) в виде:

$$a_{ik}^r = 2 \frac{\sin[\frac{\Omega_{r+1} - \Omega_r}{2}(i-k)]}{\pi(i-k)} \cos[\frac{\Omega_{r+1} + \Omega_r}{2}(i-k)] = 2a_{ik}^1 \cos[\omega_r(i-k)], \quad (3)$$

$$A_1 = \{a_{ik}^1\} = \{\sin[\frac{\Delta\Omega}{2}(i-k)] / \pi(i-k)\}, \quad (4)$$

где  $A_1$  - субполосная матрица для частотного интервала, расположенного вблизи нуля частотной оси;  $\Delta\Omega = \Omega_{r+1} - \Omega_r$ .

$\omega_r$  в (3) обозначает центральную частоту  $r$ -го частотного интервала.

$$\omega_r = (\Omega_{r+1} + \Omega_r) / 2. \quad (5)$$

Математический аппарат субполосных матриц является эффективным инструментом для решения ряда важных задач при обработке речевых сигналов. В том числе, он позволяет осуществить вычисление точных

<sup>1</sup> Жилияков Е.Г. Вариационные методы анализа и построения функций по эмпирическим данным: моногр. / Е.Г. Жилияков. – Белгород: Изд-во, 2007. – БелГУ, 2007. – 160.

значений долей энергий анализируемых отрезков речевых сигналов в заданном частотном интервале:

$$C_r = \frac{P_r}{\|\vec{x}\|^2}, r = 1, \dots, R, \quad (6)$$

$$P_r = \frac{1}{2\pi} \int_{\omega \in \Omega_r} |X(\omega)|^2 d\omega = \vec{x}^T A_r \vec{x} = \sum_{i=1}^N \sum_{k=1}^N a_{ik}^r x_k x_i, \quad (7)$$

$$X(\omega) = \sum_{k=1}^N x_k e^{-j\omega(k-1)}, \quad (8)$$

где  $P_r$  - энергия отрезка сигнала в  $r$ -ом частотном интервале;  $\vec{x} = (x_1, \dots, x_N)^T$  - отрезок (вектор) анализируемых отсчетов речевых сигналов;  $T$  - транспонирование;  $\|\vec{x}\|^2$  - энергия анализируемого отрезка сигнала.

На этой основе разработан<sup>2</sup> эффективный способ кодирования пауз, который применяется и в данной работе. Основное содержание диссертационной работы заключается в разработке способов кодирования данных, порождаемых в присутствии звуков речи.

## **Глава 2. Разработка и исследование метода и алгоритмов многополосного кодирования и декодирования речевых сообщений.**

**Раздел 2.1. Теоретические основы многополосного кодирования речевых сигналов.** Кодированию подвергаются отрезки речевых сигналов из их совокупности  $\vec{f} = (f_1, \dots, f_L)^T$ . Для простоты изложения текущий отрезок анализа обозначим в виде:

$$\vec{x} = (x_1, \dots, x_N)^T = (f_{(i-1)N+1}, \dots, f_{iN})^T, i = L/N, \quad (9)$$

Известно, что энергия отрезков речевых сигналов сосредоточена в малой доле частоты дискретизации, т.е. выполняется:

$$\|\vec{x}\|^2 = \sum_{i=1}^N x_i^2 \approx \sum_{r \in R_1} \int_{\omega \in \Omega_r} |X(\omega)|^2 d\omega / 2\pi, \quad (10)$$

где  $R_1$  - некоторое множество непересекающихся частотных интервалов вида (2), объединение которых обозначим в виде:

$$W_\Sigma = \bigcup_{r \in R_1} W_r. \quad (11)$$

В работе показано, что длина окна анализа должна быть связана с количеством частотных интервалов следующим соотношением:

$$N = 2(2R + 1) + 1. \quad (12)$$

Дается обоснование формирования многочастотных субполосных матриц (МСМ)  $A_\Sigma$ :

<sup>2</sup> Е.Г. Жилияков, Обнаружение звуков речи на фоне шумов / Е.Г. Жилияков, С.П. Белов // Научные ведомости Серия История. Политология. Экономика. Информатика. №7(126) Выпуск 22/1, 2012г.



$$A_{\Sigma} = \sum_{r \in R_1} A_r, \quad (13)$$

с элементами вида

$$a_{ik}^{\Sigma} = \sum_{r \in R_1} a_{ik}^r = \int_{\omega \in W_{\Sigma}} e^{-j\omega(i-k)} d\omega. \quad (14)$$

Введено понятие многополосной компоненты исходного отрезка сигнала

$$\bar{y} = A_{\Sigma} \bar{x}, \quad (15)$$

которая удовлетворяет следующему критерию оптимальности

$$\int_{\omega \in W_{\Sigma}} |X(\omega) - Y(\omega)|^2 d\omega / 2\pi + \int_{\omega \notin W_{\Sigma}} |Y(\omega)|^2 d\omega / 2\pi = \min. \quad (16)$$

Минимум ищется по всем трансформантам Фурье N-мерных отрезков.

Определение. Если выполняется условие

$$\bar{x}^T A_{\Sigma} \bar{x} = m \|\bar{x}\|^2, m < 1, \quad (17)$$

то компонента (15) называется информационной многополосной компонентой на уровне  $m$ .

МСМ в соответствии с (7) и (14) является симметричной и положительно определенной. Поэтому ее можно представить в виде

$$A_{\Sigma} = \sum_{i=1}^N \lambda_i^{\Sigma} \bar{q}_i^{\Sigma} \bar{q}_i^{\Sigma T}, \quad (18)$$

где выполняется

$$\lambda_i^{\Sigma} \bar{q}_i^{\Sigma} = A_{\Sigma} \bar{q}_i^{\Sigma}, \quad (19)$$

$$\lambda_i^{\Sigma} \geq \lambda_{i+1}^{\Sigma} > 0, i = 1, \dots, N-1, \quad (20)$$

$$(\bar{q}_i^{\Sigma}, \bar{q}_k^{\Sigma}) = \begin{cases} 1, & i = k \\ 0, & i \neq k \end{cases}, \quad (21)$$

здесь символ  $(\cdot)$  означает скалярное произведение.

Подстановка (18) в (15) дает представление

$$\bar{y} = \sum_{i=1}^N \lambda_i^{\Sigma} \alpha_i \bar{q}_i^{\Sigma} = QL\bar{\alpha}, \quad (22)$$

где

$$\alpha_i = (\bar{x}, \bar{q}_i^{\Sigma}); L = \text{diag}(\lambda_1^{\Sigma}, \dots, \lambda_N^{\Sigma}); Q = (\bar{q}_1^{\Sigma}, \dots, \bar{q}_N^{\Sigma}). \quad (23)$$

Определение. Замена вектора  $\bar{x}$  на вектор проекций  $\bar{\alpha}$  называется многополосным кодированием отрезков речевых сигналов, а сами  $\alpha_i$  - многополосными кодами.

В таблице 1 приведены полученные оценки квадрата относительной погрешности замены исходного отрезка многополосной информационной компонентой (в зависимости от уровня  $m$ )

$$\varepsilon^2 = \|\bar{x} - \bar{y}\|^2 / \|\bar{x}\|^2 = 1 - \sum_{i=1}^N \lambda_i (2 - \lambda_i) \alpha_i^2 / \sum_{i=1}^N \alpha_i^2. \quad (24)$$

Таблица 1 – Величина квадрата погрешности вида (24).  $N = 127, R = 31$

	$\varepsilon^2$						
звук	m=0.86	m=0.88	m=0.90	m=0.92	m=0.94	m=0.96	m=0.98
А	0,108	0,074	0,074	0,051	0,038	0,021	0,012
И	0,097	0,097	0,047	0,047	0,047	0,028	0,015
Ч	0,127	0,091	0,065	0,065	0,044	0,030	0,014
Ш	0,112	0,086	0,076	0,058	0,049	0,025	0,014

Легко заметить, что величины квадрата погрешностей сравнимы с величиной  $1 - m$ , что говорит о том, что многополосная информационная компонента адекватно отражает сосредоточенность заданной доли энергии

**Раздел 2.2 Исследование свойств собственных векторов многочастотных субполосных матриц.** В разделе основное внимание уделено исследованию частотных свойств речевых сигналов и исследованию особенностей собственных векторов МСМ. Показано, что собственное число равно доле энергии соответствующего собственного вектора, попадающей в информационный частотный интервал

$$\lambda_i^\Sigma = \int_{\omega \in W_\Sigma} |G_i^\Sigma(\omega)|^2 d\omega / 2\pi \leq \int_{-\pi}^{\pi} |G_i^\Sigma(\omega)|^2 d\omega / 2\pi = 1, \quad (25)$$

где  $G_i^\Sigma(\omega)$  – трансформанта Фурье соответствующего собственного вектора.

Отсюда следует неравенство

$$\lambda_i^\Sigma \leq 1. \quad (26)$$

Были проведены вычислительные эксперименты по оценке отношения суммарной ширины информационных частотных интервалов (11) для всех звуков русской речи:

$$\gamma = \sup W_\Sigma / \pi, \quad (27)$$

где  $\sup W_\Sigma = \Delta\Omega(\text{int}\{R_1\} - 0,5\text{Ind})$ ;

$\text{Int}\{R_1\}$  - мощность множества  $R_1$  (количество учитываемых частотных интервалов);  $\text{Ind} = \begin{cases} 0, \text{ когда интервал вблизи нуля не включен;} \\ 1, \text{ в противном случае;} \end{cases}$ .

В качестве примера, в Таблице 2 приведены усредненные результаты оценки (27) для звуков «А», «И», «Ч», «Ш».

Таблица 2 – Величина  $\gamma$  при  $N = 127, R = 31$ .

	$\gamma$						
звук	m=0.86	m=0.88	m=0.90	m=0.92	m=0.94	m=0.96	m=0.98
А	0,19	0,22	0,22	0,25	0,28	0,34	0,41
И	0,06	0,06	0,09	0,09	0,09	0,16	0,25
Ч	0,19	0,22	0,25	0,25	0,28	0,31	0,38
Ш	0,28	0,31	0,34	0,38	0,41	0,5	0,59

Результаты проведенных вычислительных экспериментов показали, что для большинства звуков русской речи выполняется неравенство  $\gamma \leq 0,25$ , и только для некоторых из них (шипящих)  $\gamma \approx 0,5 - 0,6$ .

Для каждого звука русской речи были проведены исследования по оценке количества собственных чисел МСМ, величина которых попадает в определенный интервал значений. Исследования проводились для различных значений  $R$  из соотношения (12) и различных значений параметра  $m$  из (17). В качестве примера в Таблице 3 для звука «Ш» приведено количество собственных чисел, величины которых попадают в определенный интервал значений.

Таблица 3 – Количество собственных чисел, попадающих в определенные интервалы значений  $N = 127$ ,  $R = 31$ . Звук «Ш».

Интервал значений	$m=0.86$ $\gamma = 0,28$	$m=0.88$ $\gamma = 0,31$	$m=0.90$ $\gamma = 0,34$	$m=0.92$ $\gamma = 0,38$	$m=0.94$ $\gamma = 0,41$	$m=0.96$ $\gamma = 0,50$	$m=0.98$ $\gamma = 0,59$
0 – 0.01	82	74	73	66	62	42	34
0.01 – 0.1	4	6	5	6	6	10	8
0.1-0.3	3	6	3	5	5	10	8
0.3-0.7	3	2	3	3	3	1	1
0.7-0.99	15	18	15	19	19	29	27
0.99-1	20	21	28	28	32	35	49

В таблице 4 приведены результаты вычислительных экспериментов по сопоставлению количества собственных значений больших 0.03 (параметр  $\hat{j}$ ) и их оценок на основе соотношения вида:

$$J_{\Sigma} = [4 \text{int}\{R_1\} - \text{Ind}] + 2, \quad (28)$$

где квадратная скобка означает целую часть от содержимого.

Таблица 4 – количество собственных чисел больших 0.03 и вычисленное из соотношения (28),  $N = 127$ ,  $R = 31$ .

звук	$m=0.86$		$m=0.88$		$m=0.90$		$m=0.92$		$m=0.94$		$m=0.96$		$m=0.98$	
	$\hat{j}$	$J_{\Sigma}$	$\hat{j}$	$J_{\Sigma}$	$\hat{j}$	$J_{\Sigma}$	$\hat{j}$	$J_{\Sigma}$	$\hat{j}$	$J_{\Sigma}$	$\hat{j}$	$J_{\Sigma}$	$\hat{j}$	$J_{\Sigma}$
<b>А</b>	25	26	29	30	29	30	33	34	37	38	45	46	53	54
<b>И</b>	9	10	9	10	13	14	13	14	13	14	21	22	33	34
<b>Ч</b>	25	26	30	30	33	34	33	34	37	38	41	42	48	50
<b>Ш</b>	38	38	41	42	46	46	50	50	54	54	65	66	77	78

Экспериментально было установлено, что количество собственных чисел, определяемых соотношением (28), позволяет найти количество собственных чисел, значение которых больше чем 0.03, с точностью до 1-2 значений.

В соответствии с этим, соотношение (22) для вычисления многополосной информативной компоненты примет вид

$$\bar{y} = \sum_{i=1}^{J_{\Sigma}} \lambda_i^{\Sigma} \alpha_i \bar{q}_i^{\Sigma} \quad (29)$$

и, следовательно, многополосный код состоит из  $J_{\Sigma}$  скалярных произведений вида (23).

На основе соотношения (24) были вычислены оценки погрешностей замены исходного отрезка многополосной информационной компонентой, которые приведены в Таблице 5.

Таблица 5 – Величина квадрата погрешности вида (24).  $N = 127$ ,  $R = 31$

звук	m=0.86	m=0.88	m=0.90	m=0.92	m=0.94	m=0.96	m=0.98
А	0,122	0,079	0,079	0,058	0,048	0,026	0,015
И	0,098	0,098	0,048	0,048	0,048	0,029	0,015
Ч	0,133	0,094	0,068	0,068	0,049	0,034	0,015
Ш	0,115	0,091	0,078	0,061	0,050	0,029	0,016

Можно заметить, что величина квадрата погрешности замены исходного отрезка многополосной информационной компонентой из  $J_{\Sigma}$  значений сопоставима с погрешностью, приведенной в таблице 1.

Таким образом, в соответствии с таблицей 2, на основе соотношения (29) можно сократить количество речевых данных в 4-6 раз.

**Раздел 2.3. Квантование по уровню субполосных кодов при сжатии объемов битовых представлений.** Для повышения эффективности сжатия речевых данных, целесообразно использовать процедуру квантования по уровню результатов многополосного кодирования.

Квантование по уровню имеет вид

$$\alpha_i^* = d_k, \text{ если } \alpha_i \in D_k$$

где  $d_k$  - уровни квантования;  $D_k$  - интервалы квантования.

В разделе приводятся результаты исследования трех видов квантования по уровню: равномерное квантование

$$d_{k+1} = d_k + \Delta\alpha, d_1 = \alpha_{\min} + \alpha/2, \quad (30)$$

$$\Delta\alpha = (\alpha_{\max} - \alpha_{\min}) / M. \quad (31)$$

где  $d_k$  - уровни квантования;  $M$  - заранее определяемое количество уровней.

Квантование на основе минимизации квадратической нормы погрешности, когда уровни и интервалы квантования выбираются из условия минимизации евклидовой нормы погрешности представления исходных данных квантованными:

$$\rho_{ME}^2 = \sum_{k=1}^M \sum_{i \in I_k} (\alpha_i - d_k)^2 = \min, \forall \vec{d} = (d_1, \dots, d_M)^T \quad (32)$$

и минимизации абсолютной погрешности (манхэттенская метрика)

$$\rho_{MM}^2 = \sum_{k=1}^M \sum_{i \in I_k} |\alpha_i - d_k| = \min, \forall \vec{d} = (d_1, \dots, d_M)^T, \quad (33)$$

где  $I_k$  - множество индексов квантуемой величины, когда выполняется условие

$$\alpha_i \in D_k, k = 1, M \quad (34)$$

где  $D_k$  - некоторая область возможных значений исходной величины (интервалы квантования);

В случае использования принципа (32) уровни квантования вычисляются следующим образом:

$$d_k = \sum_{i \in I_k} z_i / \text{int } I_k, \quad (35)$$

$$z_i \in D_k, \quad (36)$$

где  $z_i$  - упорядоченная по возрастанию последовательность исходных кодов, т.е.

$$z_i \in \{\alpha_k\}, k = 1, \dots, J_\Sigma, z_i < z_{i+1}, i = 1, \dots, J_\Sigma - 1. \quad (37)$$

Условию (33) удовлетворяют уровни квантования

$$d_k = z_{i_s}, \quad (38)$$

где  $z_{i_s}$  - срединное значение индекса подпоследовательности  $z_i \in D_k, i \in I_k$ .

Для восстановления многополосных информационных компонент используется представление

$$\tilde{y} = \sum_{i=1}^{J_\Sigma} \lambda_i \alpha_i^* \tilde{q}_i^\Sigma. \quad (39)$$

где  $\alpha_i^*$  - квантованные значения многополосных кодов.

С помощью вычислительных экспериментов были оценены относительные погрешности восстановления  $\tilde{y}$  при различных количествах уровней квантования. Погрешности вычислялись следующим образом: для квантования на основе (32)

$$\varepsilon_{ME} = \sqrt{\left( \frac{\|\tilde{y} - \hat{y}\|^2}{\|\tilde{y}\|^2} \right)}, \quad (40)$$

для квантования на основе (33)

$$\varepsilon_{MM} = \left( \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{\sum_{i=1}^N |y_i|} \right). \quad (41)$$

где черта сверху означает усреднение.

Для равномерного квантования погрешность оценивалась как по выражению (40), так и по (41). Погрешности усреднялись более чем по 120 экспериментам для одного и тот же звука. Результаты этих усреднений для некоторых звуков русской речи приведены в Таблице 6 ( $\varepsilon_{\text{равн}E}, \varepsilon_{\text{равн}M}$  погрешности восстановления  $\tilde{y}$  при равномерном квантовании нормированные к средней и абсолютным величинам соответственно).

Таблица 6 – Погрешность восстановления  $\hat{y}$ ,  $N = 127$ ,  $R = 31$ ,  $M = 2,4$ ,  $m = 0.92$

звук	$M = 2$				$M = 4$			
	$\varepsilon_{MM}$	$\varepsilon_{равнM}$	$\varepsilon_{ME}$	$\varepsilon_{равнE}$	$\varepsilon_{MM}$	$\varepsilon_{равнM}$	$\varepsilon_{ME}$	$\varepsilon_{равнE}$
<b>А</b>	0,09	0,09	0,22	0,23	0,06	0,06	0,16	0,16
<b>И</b>	0,04	0,05	0,16	0,18	0,03	0,03	0,12	0,13
<b>Ч</b>	0,29	0,33	0,42	0,50	0,21	0,19	0,26	0,29
<b>Ш</b>	0,21	0,24	0,36	0,42	0,15	0,14	0,23	0,25

Проведенные вычислительные эксперименты показали целесообразность использования квантования вида (32).

Было проведено исследование по оценке минимального размера квантуемой последовательности (параметр  $J_\alpha$ ), когда целесообразно вместо равномерного квантования использовать квантование, оптимальное в смысле соотношения (32) (Таблица 7).

Таблица 7 – Минимальные размеры квантуемых последовательностей, при которых целесообразно использовать оптимальное квантование вида (32)

$M$	$M = 2$	$M = 4$	$M = 8$	$M = 16$
$J_\alpha$	8	25	55	115

**Раздел 2.4 Алгоритмы многополосного кодирования/декодирования отрезков речевых сигналов.** В разделе приведено словесное описание и блок-схемы алгоритмов кодирования и декодирования речевых сообщений. Указано, какие параметры необходимо задать для начала работы алгоритма ( $R, m, \Delta\Omega, n$ ). Приводятся вычислительные эксперименты по оценке коэффициента сжатия для всех звуков русской речи.

Основные этапы алгоритма **кодирования**:

1) Считать отрезок речевого сигнала, предназначенный для текущего кодирования  $\vec{x} = (x_1, \dots, x_N)^T$ ;

2) Определить границы частотных интервалов  $\Omega_r = (r - 0.5)\Delta\Omega$ ,  $r = 1, \dots, R + 1$ ;

3) Вычислить субполосные матрицы для частотных интервалов, используя соотношения (3) и (4);

4) Вычислить доли энергии, попадающие в частотные интервалы, используя соотношение (6);

5) Определить множество  $R_1$  учитываемых частотных интервалов и зафиксировать вектор признаков многополосных информационных компонент  $\vec{maska}$ ;

6) Сформировать многочастотную субполосную матрицу, соотношение (13);

7) Вычислить собственные вектора и числа на основе определения (19);

8) Вычислить параметр  $J_\Sigma$  (28);

9) Вычислить многополосные коды вида (23);

10) Вычислить уровни квантования;

11) Определить двоичные коды уровней квантования для многополосных кодов

12) Записать закодированные данные, т.е. коды уровней квантования и служебную информацию (параметры кодирования, вектор признаков  $\vec{maska}$ , значения уровней квантования  $d_k$  и сведения о паузах).

13) Перейти к следующему отрезку речевых данных и повторить п.4 – п.13, пока не исчерпаются кодируемые данные.

**Декодирование отрезка речевых данных.** Предполагается, что при декодировании имеется заранее сформированная матрица собственных векторов первой субполосной матрицы.

1) Считать данные из файла, относящиеся к первому отрезку;

2) По кодам уровней квантования и значениям уровней  $d_k$  восстановить субполосные коды;

3) На основании вектора признаков  $\vec{maska}$  сформировать многочастотную субполосную матрицу вида (13);

4) Вычислить собственные вектора и числа на основе определения (19);

5) Вычислить параметр  $J_\Sigma$  (28);

6) Вычислить восстанавливаемый отрезок речевого сигнала  $\hat{x}_k = \sum_{i=1}^{J_\Sigma} \lambda_i \tilde{\alpha}_i^\Sigma q_{ki}^\Sigma, i = 1, J_\Sigma;$

7) Считать данные для следующего отрезка, повторить п.2 – и.7.

8) После окончания обработки всех отрезков, сформировать вектор восстановленных речевых данных;

9) Восстановить паузы на основании сохраненных сведений;

10) Записать восстановленное речевое сообщение в память.

Эффективность метода многополосного кодирования оценивалась с точки зрения коэффициента минимизации битовых представлений при 8 разрядном коде исходных отсчетов:

$$K = \frac{L * 8}{MAP + NN(R + 1) + n \sum_{i=1}^{NN} J_\Sigma(i)} \quad (42)$$

где  $L$  – общее количество отсчетов;  $MAP$  – объем бит для кодирования пауз;  $NN$  - количество обрабатываемых отрезков без пауз;  $NN(R + 1)$  – количество бит для хранения всех векторов признаков  $\vec{maska}$ ;  $J_\Sigma(i)$  - количество многополосных кодов для  $i$ -го отрезка;  $n$  – разрядность квантователя.

Было проведено более 100 вычислительных экспериментов по апробации разработанных алгоритмов. В качестве примера, в Таблице 8 приведены достигаемые показатели сжатия при обработке 20 различных речевых сигналов. Длительность речевых сигналов в среднем составляла от 45с. до 90с.

Таблица 8 – Достигаемые коэффициенты сжатия речевых данных, при  $m = 0.92$ ,  $M = 2$ .

Порядковый номер кодируемого речевого сообщения									
1	2	3	4	5	6	7	8	9	10
31,90	30,10	31,01	30,61	26,13	27,80	26,80	27,00	29,90	29,90
Порядковый номер кодируемого речевого сообщения									
11	12	13	14	15	16	17	18	19	20
27,65	27,43	26,22	28,51	27,42	26,05	30,72	28,60	26,92	26,41

Проведенные вычислительные эксперименты показали, что разработанный метод многополосного кодирования речевых сообщений на основе МСМ позволяет достигать 30 кратного сокращения объема битовых представлений исходных речевых данных. Оценка получаемого при этом качества воспроизведения восстановленных речевых сообщений приведена в 4 главе диссертации.

### Глава 3. Разработка и исследование метода и алгоритмов кодирования и декодирования речевых сообщений на основе субполосной дискретизации речевых сигналов.

**Раздел 3.1 Выделение огибающих субполосных компонент и их дискретизация.** Субполосная компонента отрезка  $\vec{x} = (x_1, \dots, x_N)^T$

$$\vec{y}_r = A_r \vec{x} \quad (43)$$

с учетом (18) может быть представлена в виде

$$\vec{y}_r = Q_1^r L_1^r Q_1^{rT} \vec{x} = Q_1^r \vec{\beta}_r, \quad (44)$$

где  $\vec{\beta}_r = L_1^r Q_1^{rT} \vec{x}$ ;  $Q_1^r = (\vec{q}_1^r, \dots, \vec{q}_J^r)$ ;  $L = \text{diag}(\lambda_1^r, \dots, \lambda_J^r)$ ;  $J$  - количество учитываемых собственных чисел.

Для компонент собственных векторов справедливо представление

$$q_{i,2k-1}^r = q_{i,k}^1 \cos[\omega_r(i-1)]; q_{i,2k}^r = q_{i,k}^1 \sin[\omega_r(i-1)], k = 1, 2, \dots, J/2, i = 1, \dots, N, \quad (45)$$

причем  $\lambda_{2k-1}^r = \lambda_{2k}^r = \lambda_k^1, k = 1, 2, \dots, J/2$ .

Поэтому компоненты вектора (44) могут быть представлены в виде

$$y_{ir} = z_{ci}^r \cos(\omega_r(i-1)) + z_{si}^r \sin(\omega_r(i-1)), \quad (46)$$

где  $z_{ci}^r = (\vec{x}, \vec{q}_{2k-1}^r)$ ;  $z_{si}^r = (\vec{x}, \vec{q}_{2k}^r)$ .

Векторы

$$\vec{z}_c^r = (z_{c1}^r, \dots, z_{cN}^r); \vec{z}_s^r = (z_{s1}^r, \dots, z_{sN}^r) \quad (47)$$

естественно называть огибающими соответствующих косинуса и синуса.

Для их выделения используются преобразования вида

$$\begin{aligned} \vec{z}_c^r &= Q^1 Q^{1T} G_c^r \vec{y}_r, \\ \vec{z}_s^r &= Q^1 Q^{1T} G_s^r \vec{y}_r, \end{aligned} \quad (48)$$

где  $G_c^r = \text{diag}(1, \cos(\omega_r), \dots, \cos(\omega_r(N-1)))$ ;  $G_s^r = \text{diag}(0, \sin(\omega_r), \dots, \sin(\omega_r(N-1)))$ ;

$Q^1 = (\vec{q}_1^1, \dots, \vec{q}_{J/2}^1)$ .



Показана возможность осуществить дискретизацию огибающих с шагом

$$N_\delta = \pi / \Delta\Omega = N/2, \quad (49)$$

т.е. оставляя только две компоненты, например, первую и  $N/2+1$ , которые и являются кодами сжимаемых отрезков речевых сигналов.

Для интерполяции используются соотношения

$$\begin{aligned} \tilde{Z}_c^r &= \frac{N}{2} Q^1 Q^{1T} \bar{w}_c^r, \\ \tilde{Z}_s^r &= \frac{N}{2} Q^1 Q^{1T} \bar{w}_s^r, \end{aligned} \quad (50)$$

где компоненты векторов  $\bar{w}_c^r$  и  $\bar{w}_s^r$  состоят из оставленных значений исходных кодов огибающих на соответствующих местах, и нулей для остальных индексов; коэффициент  $N/2$  позволяет компенсировать потерю энергии при дискретизации.

Восстановление субполосной компоненты осуществляется согласно соотношения

$$\bar{y}_r = G_c^r \tilde{Z}_c^r + G_s^r \tilde{Z}_s^r. \quad (51)$$

В свою очередь для восстановления всего отрезка используется представление

$$\hat{x} = \sum_{r=0}^R \hat{y}_r. \quad (52)$$

Ясно, что погрешности восстановления огибающих определяются погрешностями интерполяции собственных векторов матрицы  $A_1$ . Поэтому оценка погрешностей осуществлена на основе линейных комбинаций вида

$$\bar{Z} = \sum_{i=1}^{J/2} \bar{q}_i^1, \quad (53)$$

т.е. когда собственные вектора вносят одинаковый вклад.

Были проведены вычислительные эксперименты по определению комбинации отсчетов собственных векторов  $q_{i,k}^1, i=1, J/2, k=1, N$ , оставление которых приведет к минимуму квадратической погрешности результата интерполяции.

$$\varepsilon_{\text{int}} = \sqrt{\left[ \left\| \bar{z} - \bar{z}_m \right\|^2 / (J/2) \right]}. \quad (55)$$

$m = 1, 2, \dots, N/2$  - индекс первого оставляемого значения.

Результаты показали, что целесообразно в качестве начального сохраняемого индекса использовать примерно  $k = N/4$  (остальные будут определяться как  $k + N/2$ ). В Таблице 9 приведены усредненные оценки погрешности интерполяции огибающих синуса и косинуса (для различных частотных интервалов), при использовании такой комбинации отсчетов. Эксперименты проводились для 2000 различных последовательностей.

Таблица 9 – Оценки погрешности интерполяции огибающих  $\bar{z}_s^r$ ,  $\bar{z}_c^r$ .  
 $N = 127$ ,  $R = 31$ ,  $N_\partial = 63$ .

$r$	1	2	3	4	5	6	7	8	9	10
$\varepsilon_{\text{int}}(\bar{z}_s^r)$	0	0,052	0,027	0,021	0,018	0,017	0,015	0,015	0,015	0,015
$\varepsilon_{\text{int}}(\bar{z}_c^r)$	0,082	0,065	0,033	0,024	0,021	0,020	0,019	0,018	0,018	0,018

**Раздел 3.2 Квантование по уровню результатов субполосной дискретизации.** В разделе приводятся вычислительные эксперименты по оценке погрешности восстановления отрезка речевых данных по квантованным отсчетам огибающих (50). Погрешности вычислялись аналогично выражениям (40) и (41). Результаты оценки погрешности для некоторых звуков русской речи приведены в Таблице 10.

Таблица 10 – Погрешности квантования,  $N = 127$ ,  $R = 31$ ,  $N_\partial = 63$ ,  $m = 0.92$

ЗВУК	$M = 2$				$M = 4$			
	$\varepsilon_{MM}$	$\varepsilon_{\text{равн}M}$	$\varepsilon_{ME}$	$\varepsilon_{\text{равн}E}$	$\varepsilon_{MM}$	$\varepsilon_{\text{равн}M}$	$\varepsilon_{ME}$	$\varepsilon_{\text{равн}E}$
<b>А</b>	0,09	0,10	0,23	0,25	0,06	0,07	0,16	0,16
<b>И</b>	0,04	0,04	0,14	0,16	0,03	0,03	0,12	0,12
<b>Ч</b>	0,29	0,32	0,43	0,49	0,20	0,19	0,26	0,28
<b>Ш</b>	0,21	0,24	0,36	0,42	0,15	0,14	0,23	0,25

Проведенные вычислительные эксперименты показали, что квантование вида (32) имеет меньшую погрешность восстановления, чем другие методы квантования.

**Раздел 3.3 Алгоритмы субполосного кодирования/декодирования речевых сообщений на основе субполосной дискретизации речевых сигналов.** В разделе приведено словесное описание и блок-схема алгоритма кодирования и декодирования речевых сообщений. Указано, какие параметры необходимо задать для начала работы алгоритма ( $R, m, \Delta\Omega, n, N_\partial = N/2$ ). Приводятся вычислительные эксперименты по оценке коэффициента сжатия для всех звуков русской речи.

Основные этапы алгоритма **кодирования**:

- 1) Считать отрезок речевого сигнала, предназначенный для текущей обработки  $\bar{x} = (x_1, \dots, x_N)^T$ ;
- 2) Определить границы частотных интервалов  $\Omega_r = (r - 0.5)\Delta\Omega, r = 1, \dots, R + 1$ ;
- 3) Вычислить субполосные матрицы для частотных интервалов, используя соотношения (3) и (4);
- 4) Вычислить доли энергии, попадающие в частотные интервалы, используя соотношение (6);
- 5) Определить множество  $R_1$  учитываемых частотных интервалов и зафиксировать вектор признаков многополосных информационных компонент  $\vec{maska}$ ;
- 6) Вычислить субполосные компоненты, используя соотношение (46);

- 7) Вычислить огибающие косинуса и синуса для этих компонент, используя (48);
- 8) Продискретизировать огибающие с шагом  $N_\partial = N/2$ ;
- 9) Вычислить уровни квантования;
- 10) Определить двоичные коды уровней квантования для оставшихся отсчетов;
- 11) Записать закодированные данные, т.е. коды уровней квантования и служебную информацию (параметры кодирования, вектор признаков  $\vec{maska}$ , значения уровней квантования  $d_k$  и сведения о паузах).
- 12) Перейти к следующему отрезку речевых данных и повторить п.4 – п.12, пока не исчерпаются кодируемые данные.

**Декодирование отрезка речевых данных.** Предполагается, что при декодировании имеется заранее сформированная матрица собственных векторов первой субполосной матрицы.

- 1) Считать данные из файла, относящиеся к первому отрезку;
- 2) По кодам уровней квантования и значениям уровней  $d_k$  восстановить данные;
- 3) Осуществить интерполяцию огибающих согласно выражению (50);
- 4) Восстановить субполосную компоненту, используя выражение (51);
- 5) Восстановить отрезок речевых данных, используя выражение (52);
- 6) Считать данные для следующего отрезка, повторить п.2 – п.6.
- 7) После окончания обработки всех отрезков, сформировать вектор восстановленных речевых данных;
- 8) Восстановить паузы на основании сохраненных сведений;
- 9) Записать восстановленное речевое сообщение в память.

Эффективность метода оценивалась с точки зрения коэффициента минимизации битовых представлений:

$$K = \frac{L * 8}{MAP + NN(R + 1) + n \sum_{i=1}^{NN} 2K_m \left[ \frac{N}{N\partial} \right]}, \quad (54)$$

где  $L$  – общее количество отсчетов;  $MAP$  – сведения о паузах;  $NN$  – количество обрабатываемых отрезков без пауз;  $NN(R + 1)$  – количество бит для хранения всех векторов признаков  $\vec{maska}$ ;  $2K_m \left[ \frac{N}{N\partial} \right]$  – количество отсчетов огибающих косинуса и синуса после дискретизации;  $[]$  – целая часть от содержимого;  $n$  – разрядность квантователя.

Было проведено более 100 вычислительных экспериментов по апробации разработанных алгоритмов. В качестве примера, в Таблице 11 приведены достигаемые показатели сжатия при обработке 20 различных речевых сигналов. Длительность речевых сигналов в среднем составляла от 45с. до 90с.

Таблица 11 – Достижимые коэффициенты сжатия речевых данных, при  $m = 0.92$ ,  $M = 2$ ,  $N_{\delta} = N/2$

	Порядковый номер кодируемого речевого сообщения									
	1	2	3	4	5	6	7	8	9	10
$M = 2$	41,47	39,13	40,31	39,79	33,97	36,14	34,84	35,10	38,87	38,87
	Порядковый номер кодируемого речевого сообщения									
	11	12	13	14	15	16	17	18	19	20
$M = 2$	35,95	35,66	34,09	37,06	35,65	33,87	39,94	37,18	35,00	34,33

Проведенные вычислительные эксперименты показали, что разработанный метод кодирования на основе субполосной дискретизации речевых сигналов позволяет достигать 35 кратное сокращение объема битовых представлений исходных речевых данных. Оценка получаемого при этом качества воспроизведения восстановленных речевых сообщений приведена в 4 главе диссертации.

#### Глава 4. Разработка программных реализаций алгоритмов кодирования речевых сигналов.

**Раздел 4.1 Разработка рекомендаций по использованию разработанных методов кодирования речевых данных.** В разделе проведены вычислительные эксперименты по оценке качества восстановленных речевых сообщений. Оцениванию подвергались восстановленные сообщения, которые были закодированы при различных параметрах кодирования. Для оценивания качества использовалась специальная форма, в которую экспертная группа вносила оценки (от 1 до 5) для таких показателей как: уровень разборчивости, комфортность прослушивания, узнаваемость диктора. Также они могли отметить наличие каких-либо искажений в сигнале (зашумление, треск и т.д.). По результатам усреднения выставлялась итоговая оценка.

По результатам оценивания были сформированы три рекомендуемых типа параметров кодирования (Таблица 12 МК – метод многополосного кодирования, СД – метод на основе субполосной дискретизации).

Таблица 12. Описание рекомендуемых параметров кодирования.

	Параметры кодирования		Коэффициент сжатия		Оценка качества	
	МК	СД	МК	СД	МК	СД
«Максимальное сжатие»	$m=0.92$ , $M=2$	$m=0.92$ , $M=2$ , $N_{\delta}=N/2$	26-31	30-40	3,5-3,9	3,2-3,5
«Среднее сжатие»	$m=0.94$ , $M=4$	$m=0.94$ , $M=4$ , $N_{\delta}=N/2$	16-24	14-25	3,9-4,2	3,9-4,3
«Низкое сжатие»	$m=0.96$ , $M=8$	$m=0.96$ , $M=8$ , $N_{\delta}=N/3$	11-14	10-12	4,3-4,6	4,3-4,6

Отсюда следует, что для достижения лучшего качества воспроизведения речевых сообщений целесообразно использовать метод МК, тогда как метод СД позволяет достичь больших степеней сжатия.

Были проведены вычислительные эксперименты по сравнению разработанных алгоритмов кодирования с известными современными аналогами (Таблицы 13 - 14).

Таблица 13 Сравнение с современными аналогами

	МК	СД	MP3	GSM 6.10	DSP Group TrueSpeech	AAC
Коэффициент сжатия	12,82	10,91	5,84	7,24	8,13	9,87
Оценка качества	4,3	4,4	4,7	4,6	4,3	4,3

Таблица 14 Сравнение с современными аналогами

	МК	СД	Sound Squeezer	Speex
Коэффициент сжатия	29,81	35,62	26,56	22,47
Оценка качества	3,58	3,41	3,6	3,7

В качестве вывода к разделу можно указать, что разработанные методы и алгоритмы кодирования превосходят современные аналоги по степени сжатия, при этом достигается сопоставимый уровень качества воспроизведения.

**Раздел 4.2 Программно-алгоритмическая поддержка разработанных алгоритмов кодирования речевых данных.** Приводится описание прототипа программной реализации разработанных алгоритмов кодирования и декодирования речевых сигналов.

**Раздел 4.3 Разработка структуры пакета для хранения и передачи закодированных речевых данных.** Содержание раздела посвящено разработке структуры пакета речевых данных, закодированных на основе созданных процедур кодирования. Для каждого типа данных отмечено количество бит, необходимое для хранения. В качестве примера, на рисунке 1 приведена структура пакета закодированных речевых данных для метода многополосного кодирования.

Служебная информация					Данные	
Информация о паузах			Параметры кодирования	Вектор признаков	Значения уровней квантования	Закодированные речевые данные
Размер карты	Начало паузы	Количество отрезков паузы	4 вида возможных настроек	<i>mask</i>	$d_k$	Коды уровней квантования
8, бит	8, бит	8, бит	2, бита	R, бит	$2^n * 8$ , бит	$n * J_2$ , бит

Рисунок 1 – структура пакета речевых данных для метода многополосного кодирования

**В заключении** сформулированы основные результаты и выводы диссертационной работы.

### ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ:

1. Разработаны методы и алгоритмы кодирования речевых сообщений на основе субполосного анализа и синтеза речевых сигналов, позволяющие минимизировать объемы битовых представлений речевых

данных при условии сохранения заданного уровня информативности восстанавливаемых речевых сообщений, включая:

1.1. Метод и алгоритм выделения многополосных информационных компонент речевых сигналов, на основе определения частотных интервалов минимальной суммарной ширины, содержащих заданную долю энергии анализируемого отрезка.

1.2. Метод и алгоритм многополосного кодирования и декодирования речевых сообщений на основе оптимальной аппроксимации трансформант Фурье в адаптивно определяемых частотных интервалах, позволяющий сократить объем битовых представлений исходных речевых данных до 30 раз, при сохранении разборчивости воспроизводимой речи и ее тембральных характеристик.

1.3. Метод и алгоритм кодирования и декодирования речевых сообщений на основе субполосной дискретизации огибающих информационных субполосных компонент речевых сигналов, позволяющий сократить объем битовых представлений исходных речевых данных до 35 раз, при сохранении разборчивости воспроизводимой речи.

2. Разработана структура пакета речевых данных, закодированных на основе созданных процедур кодирования. Она содержит закодированные речевые данные, а также служебную информацию, которая необходима для успешного декодирования речевых сообщений. Для всех содержащихся в пакете сведений указано необходимое количество бит, которым они кодируются.

3. Создан прототип программной поддержки разработанных методов и алгоритмов и сформулированы рекомендации по его использованию. В рекомендациях содержатся сведения для пользователя, которые характеризуют параметры кодирования и пояснения, какой результат может быть получен при выборе тех или иных параметров.

4. Результаты проведенных вычислительных экспериментов, подтверждают обоснованность и достоверность выводов, а также эффективность разработанных методов субполосного кодирования речевой информации.

## **ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ**

### *Статьи в журналах из списка ВАК*

1. Болдышев А.В. Оптимальное двоичное кодирование уровней речевых данных / Жиляков Е.Г., Белов С.П., Прохоренко Е.И., Болдышев А.В., Фирсова А.А. // «Вопросы радиоэлектроники», серия ЭВТ. Выпуск №1, Москва, 2013. с. 110-115.

2. Болдышев А.В., Алгоритм сжатия речевых данных на основе двумерной обработки данных. / Е.Г. Жиляков, Е.И. Прохоренко, А.В. Болдышев // Журнал «Вопросы Радиоэлектроники», серия ЭВТ. Выпуск №1, Москва, 2012. С.27-33

3. Болдышев А.В. Сжатие речевых данных на основе прореживания результатов оптимальной линейной частотной фильтрации. /Е.Г. Жиляков, А.В. Болдышев, Н.А. Чеканов // Научные Ведомости БелГУ серия История. Политология. Экономика. Информатика. Белгород, №7 (126) 2012, Выпуск 22/1 с.162-167

4. Болдышев А.В. Метод сжатия речевых данных на основе составной субполосной матрицы / Е.И. Прохоренко, А.В. Болдышев, А.В. Эсауленко // Журнал «Вопросы Радиоэлектроники», серия ЭВТ Выпуск №1, Москва 2011. с. 60-72.

5. Болдышев А.В. Метод сжатия речевых данных на основе оптимального субполосного преобразования по составным частотным интервалам / А.В. Болдышев // Научные Ведомости БелГУ серия История. Политология. Экономика. Информатика. Белгород, 2011. № 1 (96). Выпуск 17/1 с.217-222

6. Болдышев А.В. Сегментация речевых сигналов на основе анализа распределения энергии по частотным интервалам / Е. Г. Жиляков, Е. И. Прохоренко, А. В. Болдышев, А. А. Фирсова, М. В. Фатова // Научные Ведомости БелГУ серия История. Политология. Экономика. Информатика. Белгород, №7 (102) 2011, Выпуск 18/1 с.187-196

7. Болдышев А.В. Об избирательном воздействии на частотные компоненты речевых сигналов в задаче сжатия / А.В. Болдышев // Научные Ведомости БелГУ серия История. Политология. Экономика. Информатика. Белгород, 2010. № 13 (84). Выпуск 15/1

8. Болдышев А.В. Метод оптимального субполосного преобразования в задаче сжатия речевых данных / Е.И. Прохоренко, А.В. Болдышев, А.А. Фирсова, А.В. Эсауленко // Журнал «Вопросы Радиоэлектроники», серия ЭВТ. Выпуск №1 Москва 2010. с. 49-55.

*Публикации в научных журналах и сборниках трудов научных конференций*

9. Болдышев А.В. Сжатие речевых данных на основе прореживания результатов оптимальной линейной частотной фильтрации / Е. Г. Жиляков, А. В. Болдышев // Информационные технологии в науке, образовании и производстве: сб. трудов V междунар. науч-практ конф. – г. Орел, 2012 г.

10. Болдышев А.В. Сжатие речевых данных на основе прореживания результатов оптимальной линейной частотной фильтрации /Е.Г. Жиляков, А.В. Болдышев // Прикладная математика, управление и информатика: сб. трудов междунар. молодеж. конф. – Белгород 2012г. с. 364-368

11. Болдышев А.В. Избирательное воздействие на частотные компоненты речевых сигналов в задаче сжатия /А.В. Болдышев, Е.И. Прохоренко, Е.Ю. Гарькавая // Цифровая обработка сигналов и её применение - DSPA: сб. трудов. 13 междунар. конф. – Москва, 2011 г. с 171-175.

12. Болдышев А.В. Об избирательном преобразовании частотных компонент речевых сигналов в задаче сжатия / Е.Г. Жиляков, А.В. Болдышев,

А.А. Фирсова, А.В. Курлов А.В. Эсауленко // MicroCAD-2011: сб. трудов. междунар. конф. – Харьков, Украина 2011г. с 37-44.

13. Болдышев А.В. Эффективное использование ресурсов информационно-телекоммуникационных систем на основе нового метода сжатия речевых данных / Е.Г. Жиляков, А.В. Болдышев, А.А. Фирсова // Медико-экологические информационные технологии: сб. трудов. 14 междунар. науч.-техн. конф. – Курск 2011 г. с. 63-71

14. Болдышев А.В. Сжатие речевых данных на основе прореживания результатов линейной оптимальной фильтрации / А.В. Болдышев, Е.И. Прохоренко // Компьютерные науки и технологии: сб. труд. 2 междунар. науч.-техн. конф. – Белгород: ООО «ГиК», 2011. с. 548-553.

15. Болдышев А.В. Об использовании особенностей распределения энергии звуков русской речи / Е.Г. Жиляков, А.В. Болдышев, А.А. Фирсова // Инновационные подходы к применению информационных технологий в профессиональной деятельности: сб. науч. труд. 2-й междунар. науч.-практ. Интернет конф. – Белгород, 2010. с. 323-327

16. Болдышев А.В. Модели распределения энергии звуков русской речи на основе частотных представлений / Е.Г. Жиляков, А.В. Болдышев, А.А. Фирсова // Математические методы в технике и технологиях (ММТТ–23): сб. трудов XXIII междунар. науч. конф. – Саратов, 2010. с. 236-239

#### *Программы для ЭВМ*

17. Свидетельство о государственной регистрации программы для ЭВМ № 2010616593 «Программная система сжатия/восстановления речевых сообщений на основе нового метода субполосного преобразования и квантования по уровню», Жиляков Е.Г., Прохоренко Е.И., Болдышев А.В., Фирсова А.А., Эсауленко А.В. от 5 августа 2010г.

18. Свидетельство о государственной регистрации программы для ЭВМ № 2011616998 «Программная система фильтрации речевых сигналов на основе составных субполосных матриц», Жиляков Е.Г., Прохоренко Е.И., Болдышев А.В., Фирсова А.А., Курлов А.В. от 08.09.2011

19. Свидетельство о государственной регистрации программы для ЭВМ № 2011617002 «Сжатие/восстановление речевых данных на основе составного субполосного анализа/синтеза», Жиляков Е.Г., Прохоренко Е.И., Болдышев А.В., Фирсова А.А., Фатова М.В. от 08.09.2011

20. Свидетельство о государственной регистрации программы для ЭВМ № 2011617801 Программная система сжатия речевых данных за счет обнаружения и кодирования пауз», Жиляков Е.Г., Прохоренко Е.И., Болдышев А.В. от 17.08.2011

Подписано в печать 22.02.2013. Гарнитура Times New Roman

Формат 60×84/16.Усл. п. л. 1,0. Тираж 100 экз. Заказ 40.

Оригинал-макет подготовлен и тиражирован в ИД «Белгород» НИУ «БелГУ»  
308015, г. Белгород, ул. Победы, д.85