

Федеральное государственное казенное военное
образовательное учреждение высшего образования
Академия Федеральной службы охраны Российской Федерации

На правах рукописи



Афанасьев Андрей Алексеевич

МОДЕЛИ И МЕТОДЫ АНАЛИЗА И ОБРАБОТКИ
РЕЧЕВОГО СИГНАЛА В СИСТЕМАХ СВЯЗИ

Специальность 05.12.13 – Системы, сети и устройства телекоммуникаций

Диссертация на соискание ученой степени
доктора технических наук

Научный консультант: доктор технических наук
доцент Левин Евгений Калманович

Орёл – 2018

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	5
1 СУЩЕСТВУЮЩИЕ ПРОБЛЕМЫ АНАЛИЗА И ОБРАБОТКИ РЕЧЕВОГО СИГНАЛА В СИСТЕМАХ СВЯЗИ.....	15
1.1 Вводные замечания	15
1.2 Анализ методов обработки речевого сигнала для задач связи.....	15
1.3 Требования, предъявляемые к системам обработки речевого сигнала	26
1.4 Структуризация проблем обработки речевого сигнала	29
1.5 Постановка задач диссертационного исследования	44
1.6 Выводы по первому разделу	46
2 РАЗРАБОТКА МЕТОДА ВЫДЕЛЕНИЯ СЕГМЕНТОВ АНАЛИЗА ПРИ МОДЕЛИРОВАНИИ ПРОЦЕССОВ ОБРАБОТКИ РЕЧЕВОГО СИГНАЛА В СИСТЕМАХ СВЯЗИ.....	48
2.1 Вводные замечания	48
2.2 Обработка речевого сигнала на основе структурно-параметрической адаптации	48
2.3 Метод деления речевого сигнала на сегменты различной длительности .	65
2.4 Определение параметров синтезирующего фильтра при кодировании речевого сигнала.....	78
2.5 Оценка и коррекция линейных спектральных частот на однородных сегментах анализа при обработке речевого сигнала	102
2.6 Выводы по второму разделу	112
3 ОБРАБОТКА РЕЧЕВОГО СИГНАЛА С УЧЕТОМ ВЗАИМОЗАВИСИМОСТИ ЭЛЕМЕНТОВ ЕГО ДЕКОМПОЗИЦИИ	113
3.1 Вводные замечания	113
3.2 Выявление зависимости элементов декомпозиции речевого сигнала....	113
3.3 Модель системы обработки речевого сигнала, учитывающая зависимость элементов его декомпозиции	116
3.4 Метод обработки речевого сигнала с учетом зависимости элементов его декомпозиции	127
3.5 Выводы по третьему разделу	153

4 МЕТОД ПОДАВЛЕНИЯ ПОМЕХ В РЕЧЕВОМ СИГНАЛЕ НА ОСНОВЕ ПРИМЕНЕНИЯ ПОЛИСПЕКТРАЛЬНОГО АНАЛИЗА	155
4.1 Вводные замечания	155
4.2 Обработка речевого сигнала при подавлении помех	155
4.3 Математическая модель системы фильтрации речевого сигнала на основе применения полиспектрального анализа.....	171
4.4 Выделение сегментов паузы и оценка спектральных характеристик помех.....	190
4.5 Выводы по четвертому разделу	205
5 РЕАЛИЗАЦИЯ ПРОЦЕДУР ПОДАВЛЕНИЯ ПОМЕХ В РЕЧЕВОМ СИГНАЛЕ	206
5.1 Вводные замечания	206
5.2 Метод адаптивной фильтрации речевого сигнала в условиях действия акустических помех	206
5.3 Критерий оценки качества речевого сигнала после его обработки.....	220
5.4 Оценка качества речевого сигнала при подавлении помех согласно предлагаемым техническим решениям.....	226
5.5 Выводы по пятому разделу	240
6 РАЗРАБОТКА АЛГОРИТМОВ ОБРАБОТКИ РЕЧЕВОГО СИГНАЛА В СИСТЕМАХ СВЯЗИ.....	241
6.1 Вводные замечания	241
6.2 Определение характеристик сегмента анализа речевого сигнала.....	241
6.3 Классификация сегментов речевого сигнала	245
6.4 Определение длительности сегмента анализа речевого сигнала	250
6.5 Определение взаимных связей элементов декомпозиции речевого сигнала.....	260
6.5.1 Процедура векторного квантования элементов декомпозиции речевого сигнала.....	260
6.5.2 Алгоритм установления взаимосвязей элементов декомпозиции речевого сигнала.....	271
6.5.3 Кодирование (декодирование) речевого сигнала с учетом зависимости	

элементов декомпозиции речевого сигнала	275
6.6 Особенности реализации предложенных методов обработки речевого сигнала в системах телекоммуникаций	283
6.7 Выводы по шестому разделу.....	292
ЗАКЛЮЧЕНИЕ	294
СПИСОК СОКРАЩЕНИЙ.....	297
СПИСОК ЛИТЕРАТУРЫ.....	299
ПРИЛОЖЕНИЕ 1 Особенности и условия формирования тестовых фраз для проверки разработанных методов и алгоритмов	324
ПРИЛОЖЕНИЕ 2 Акты внедрения	326

ВВЕДЕНИЕ

Актуальность темы исследования и степень ее разработанности. Одной из тенденций развития инфокоммуникационных систем в мире является переход на методы цифровой обработки сигналов (ЦОС), которые используются в различных системах связи, в том числе в мобильных устройствах, обеспечивающих повышение качества обслуживания абонентов и расширение перечня инфокоммуникационных услуг [59, 73, 74].

Преимущества ЦОС с точки зрения помехоустойчивости и качества воспроизведения общеизвестны. Однако без использования процедур сжатия и дополнительных методов представления данных требуется значительное увеличение пропускной способности каналов связи (КС), что ведет к росту стоимости их аренды и обслуживания.

Разработка эффективных способов эффективного кодирования видео- и звуковой информации является предпосылкой более эффективного использования КС, что выражается в высвобождении значительной части трафика и дает возможность обеспечить потребителям использование дополнительных видов информационных услуг.

Рост объема информации, передаваемой по сетям электросвязи, сопровождается сохранением значительной части трафика речевых сообщений. Телефонный трафик до сих пор остается одним из основных, так как речь является наиболее естественной формой человеческого общения. На сегодняшний день множество инфокоммуникационных служб используют в своем оборудовании средства цифровой обработки речевого сигнала (РС). Наиболее бурно развивающейся областью цифровой обработки РС являются системы мобильной радиосвязи (сотовой, транкинговой, спутниковой, радио), системы Интернет-телефонии, специальные системы связи и идентификации дикторов и речевого управления [1].

Разрабатываются новые высокоэффективные алгоритмы цифровой обработки РС различного назначения, возрастает количество приложений таких алгоритмов в самых разнообразных системах связи. Появляются новые способы обработки РС, которые позволяют сократить объем передаваемых данных и повысить качественные показатели функционирования систем и, благодаря этому, снизить цены на оборудование и услуги связи, наряду с удовлетворением требований пользователей таких систем [59, 78, 106, 163, 169, 180].

Достаточно большое количество работ отечественных и зарубежных авторов посвящено созданию систем обработки РС, наиболее фундаментальными из них являются труды Михайлова В. Г., Сапожкова М. А., Назарова М. В., Прохорова Ю. Н., Ланнэ А. А., Грэя А. Х., Маркела Дж. Д., Рабинера Л. Р., Ситняковского И. В., Шафера Р. В., Шалимова И. А., Петровского А. А., Ковалгина Ю. А., Итакуры Ф., Рейдера Ч., Макхоула Дж и других. Хронология развития теории и методов обработки РС представлена на рисунке 1.

40-е-50-е годы XX века. А.А.Пирогов, М.А.Сапожков, Д.Фланаган, Г.Фант. Исследование свойств, характеристик, моделей, способов представления речевого сигнала, разработка низкоскоростных вокодеров
60-е-70-е годы XX века. Ланнэ А.А., Шафер Р.В., Рабинер Л.Р., Б.Голд, Михайлов В.Г., Вемян Г.В., Дж.Д. Маркел, А.Х.Грэй, Д. Раймд, Ф.Итакура, С.Саито, Б.С.Атол Разработка моделей синтеза и анализа речевого сигнала, систем низкоскоростного кодирования речи и шумоподавления, способов выделения характеристических параметров речи, систем оценки качества кодирования речи
окончание XX века – начало XXI века Улахович Д.А., Прохоров Ю.Н., Коротаев Г.А., М.Р.Шредер, Д.Серено, А.Гершо, Шалимов И.А., Шелухин О.И., Калинин Ю. К., Петровский А.А. Разработка вычислительно сложных систем низкоскоростного кодирования речи, основанных на различных подходах к методу линейного предсказания, гибридным способам кодирования, использование психоакустических особенностей слуха человека, различных вариантов векторного квантования параметров и многомерной обработки.

Рисунок 1 – Развитие теории и методов обработки речевого сигнала

Тем не менее, необходимость исследования и разработки методов и алгоритмов обработки РС, повышающих качество синтезированной речи по показателям естественности ее звучания и узнаваемости абонента при сохранении требуемой скорости передачи, а также снижение средней скорости передачи при сохранении достоверности передаваемой информации остается актуальной и важной как с научной, так и с практической точек зрения задач.

С развитием вычислительной техники появилась возможность применения более совершенных, и, соответственно, более сложных алгоритмов обработки РС с целью снижения скорости потока данных при сохранении качества синтезированного сигнала. Их разработка требует более детального анализа существующих методов обработки РС.

Анализ технических решений и способов организации речевого обмена показывает, что далеко не полностью реализуются потенциальные возможности комплексов обработки РС [122, 148, 154]. При этом используемые в настоящее время технологии анализа и обработки РС, основанные на его аналитическом описании, не в полной мере учитывают некоторые значимые особенности формирования речевого сигнала и потенциальные возможности по совершенствованию применяемых методов.

Существующие потребности в повышении эффективности использования пропускной способности сетевых ресурсов, а также необходимость учета новых особенностей, обусловленных тенденциями развития современных инфокоммуникационных систем, определяют актуальность научного поиска в области разработки новых и совершенствования существующих методов и алгоритмов обработки РС [1, 3, 62].

Решение этой задачи связано с созданием эффективных речепреобразующих устройств (РПУ), кодеков РС (КРС) – устройств преобразования РС в цифровую форму на передающей стороне и восстановления его аналоговой формы на приемной стороне. Следует отметить, что КРС является одним из важнейших элементов системы связи при передаче по ним речевой информации, так как показатели про-

цедур обработки РС определяют потенциально достижимую верхнюю границу качества речевых услуг, предоставляемых на основе цифровых телекоммуникационных систем.

В связи с изложенным, **научная проблема** диссертационного исследования формулируется следующим образом. Снижение скорости потоков данных РС при обеспечении требуемого его качества в условиях действия акустических помех обуславливает необходимость разработки моделей и методов анализа и обработки РС, учитывающих в большей степени особенности формирования РС и технологии его представления в системах телекоммуникаций.

Целью диссертационной работы является разработка моделей и методов обработки РС в условиях действия акустических помех, обеспечивающих снижение скорости потока данных при заданном качестве РС за счет более точного учета особенностей его формирования при существующих технологиях представления в системах телекоммуникаций.

Таким образом, **объектом исследования** является РС, подвергаемый обработке в системах связи.

В качестве **предмета исследования** выступают модели, методы и алгоритмы обработки РС, а также способы их реализации в системах связи.

Научная новизна исследования обусловлена тем, что в нем:

1) впервые предложены:

- комплексное техническое решение по формированию методологии понижения скорости кодирования РС при его передаче с сохранением качественных показателей синтезированной речи в условиях воздействия акустических помех;

- метод обработки речевого сигнала, основанный на учете зависимости элементов декомпозиции РС при линейном предсказании (ЛП), позволяющий уменьшать среднюю скорость передачи в канале связи при сохранении качественных показателей синтезированной речи;

- технологии адаптивной цифровой фильтрации речевого сигнала в условиях акустического зашумления, отличающиеся применением полиспектрального анализа и позволяющие осуществлять качественное шумоподавление при низких отношениях сигнал-шум;

2) развиты теоретические принципы применения моделей обработки РС со структурно-параметрической адаптацией, на основе которых разработаны технические решения по выделению сегментов анализа различной длительности на однородных участках речевого сигнала;

3) разработан новый подход к выделению пауз в РС при наличии акустического зашумления, основанный на применении технологий полиспектрального анализа;

4) проведено обоснование новых моделей, методов и алгоритмов анализа и обработки РС с учетом принятых технологий его кодирования, учитывающих основы речеобразования, которые могут быть использованы для совершенствования устройств обработки РС, функционирующих в системах связи.

Теоретическая ценность диссертационного исследования определяется развитием методов анализа РС, разработкой моделей и методов его обработки в системах связи в различных условиях функционирования с учетом ограничений на скорость передачи в КС.

Практическая значимость диссертационного исследования заключается в использовании разработанных и запатентованных методов и алгоритмов обработки РС, которые экспериментально проверены с использованием имитационного моделирования и обеспечивают:

- понижение средней скорости передачи в канале связи при сохранении качества РС;

- функционирование систем обработки РС в условиях шумовых акустических воздействий, при этом осуществлять фильтрацию в задаче шумоподавления со средним снижением показателя *Modified Bark Spectral Distortion* (MBSD) от 4,20 до 2.88 (среднее повышение субъективной оценки от 0,87 до 1,22 балла) в диапазоне входных значений отношения сигнал-шум (ОСШ) от 15 до -5 дБ.

Внедрение и использование разработанных технических решений дает возможность внести значительный вклад в хозяйственно-экономическое развитие и обороноспособность страны.

Обоснованность и достоверность научных положений, основных выводов и результатов диссертации обеспечивается за счет комплексного анализа состояния предметной области исследования, непротиворечивости теоретических выводов с результатами экспериментальной проверки предложенных моделей на основе имитационного моделирования, а также апробацией основных теоретических положений диссертации в печатных трудах и докладах на научных и научно-практических конференциях различных уровней.

Методы исследования. Для достижения цели исследования в работе использованы методы системного анализа, декомпозиции и агрегирования, статистического и корреляционного анализа, исследования операций, теорий информации, множеств и матриц, теории вероятностей и математической статистики, методы многокритериальной оптимизации, теории кодирования и ЦОС, аналитическое и имитационное моделирование.

На защиту выносятся следующие **положения диссертационного исследования:**

1. Структурно-параметрическая адаптация, используемая в модели кодирования РС в виде свертки параметров, описывающих передаточную функцию ограниченного порядка голосового тракта и сигнала возбуждения, отличающаяся от известных учетом особенностей формирования РС и технологий его кодирования, позволяет разработать методологию снижения средней скорости передачи данных в КС.

2. Учет зависимости элементов декомпозиции РС при построении модели его обработки при низкоскоростном кодировании на основе метода ЛП, позволяет осуществить снижение мощности пространства представления сигналов возбуждения.

3. Применение полиспектрального анализа при моделировании системы фильтрации РС в задаче шумоподавления позволяет разработать методы шумоподавления в РС для диапазона значений ОСШ от 15 до -5 дБ.

4. Метод выделения сегментов анализа РС различной длительности на однородных участках РС, позволяющий осуществлять его синтез при фиксированных значениях параметров передаточной функции голосового тракта.

5. Метод обработки РС, основанный на учете зависимости элементов его декомпозиции при ЛП, позволяющий уменьшать среднюю скорость передачи данных в КС при сохранении качественных показателей синтезированного РС.

6. Метод адаптивной цифровой фильтрации РС в условиях акустического зашумления, отличающийся применением технологий полиспектрального анализа и позволяющий осуществлять качественное шумоподавление при отношениях сигнал-шум до $0 \div -5$ дБ.

7. Алгоритмы анализа и обработки РС, позволяющие повысить характеристики функционирования телекоммуникационных устройств в системах инфокоммуникаций в условиях наличия акустических шумов, отличающиеся от известных учетом особенностей формирования и обработки РС, применением результатов полиспектрального анализа с целью шумоподавления, а также учетом взаимозависимости элементов декомпозиции РС при линейном предсказании, устраняющие противоречия между особенностями формирования РС и существующими в современных системах связи технологиями обработки РС.

Апробация результатов работы. Результаты диссертационного исследования обсуждались и докладывались в период с 2001 по 2017 годы на:

1) международной научной конференции «Речь и компьютер». SPESOM'2001: Москва, МГЛУ [192];

2) XI, XVI международных научных конференциях «Информатизация и информационная безопасность правоохранительных органов». Москва, Академия УМВД, 2002, 2007 [19, 21, 29];

3) 4-й, 5-й, 6-й, 7-й, 12-й, 13-й, 15-й, 16-й международных конференциях и выставках "Цифровая обработка сигналов и ее применение". Москва: ИПУ РАН, 2002, 2003, 2004, 2005, 2010, 2011, 2013, 2014 [20, 22, 24, 25, 43, 44, 51, 52];

4) международной научной конференции к 95-летию академика В.А. Котельникова «Современная радиоэлектроника в ретроспективе идей В.А. Котельникова». – Москва: МЭИ, 2003 [23].

5) 10, 11, 13, 15, 16 всероссийских научно-технических конференциях молодых ученых и специалистов «Новые информационные технологии в научных исследованиях и образовании». – Рязань: РГРТУ, 2005, 2006, 2008, 2010, 2011 [26,27,28,35,36,37].

6) 32-й, 33-й Всероссийских научно-практических конференциях "Сети, системы связи и телекоммуникации". – Рязань: РВВКУС, 2007 [30, 31, 33] .

7) всероссийской научной конференции «Проектирование научных и инженерных приложений в среде MATLAB». – Санкт-Петербург: СПбГУ, 2007 [32].

8) 1-й всероссийской научно-технической конференции «Современные информационные технологии в деятельности органов государственной власти – Информтех 2008». – Курск: КГТУ, 2008 [34].

9) 8-й Всероссийской научно-технической Интернет-конференции. 27-28 сентября 2011 г. –Тула: ТГУ, 2011 [38, 39].

10) 3-й международной Интернет-конференции молодых ученых, аспирантов и студентов «Инновационные технологии: теория, инструменты, практика». – Пермь: ПНИПУ, 2012 [40].

11) 2-й региональной научно-практической конференции "Многоядерные процессоры, параллельное программирование, системы обработки сигналов", 2012 – Барнаул, 2012 [18].

12) 2-й международной заочной научно-технической конференции «Информационные технологии. Радиоэлектроника. Телекоммуникации (ITRT-2012)». – Тольятти: Поволжский ГУ сервиса, 2012 [41].

13) всероссийской научно-технической конференции, посвященной 117-й годовщине дня радио «Современные проблемы радиоэлектроники». – Красноярск: Сибирский ФУ, 2012 [42].

14) 11-й международной научно-технической конференции «INTERMATIC–2012». – Москва, МГТУ МИРЭА – ИРЭ РАН, 2012 [47].

15) 55-й Всероссийской научной конференции МФТИ "Проблемы фундаментальных и прикладных естественных и технических наук в современном информационном обществе". – Москва: МФТИ, 2012 [45].

16) 2-й Международной научно-практической конференции студентов и молодых ученых "Молодежь и наука: модернизация и инновационное развитие страны". – Пенза, ПенГУ, 2012 [46].

17) 2-й Международной научно-практической конференции "Инновационные технологии в производстве, науке и образовании" – Грозный, ГГНТУ, 2012 [49].

18) Конгрессе молодых ученых, Санкт-Петербург: НИУ ИТМО, 2012 [48].

19) III Всероссийской научно-технической конференции «Системы связи и радионавигации», г. Красноярск : АО «НПП «Радиосвязь», 2016 [55].

20) 16-й всероссийской с международным участием научно-технической конференции «Современные проблемы радиоэлектроники». – Красноярск: Сибирский ФУ, 2013 [50].

21) 13-й международной молодежной научно-технической конференции «Будущее технической науки». – Нижний Новгород: НГТУ, 2014 [53].

22) Международной научно-технической конференции "Информационные системы и технологии". – Нижний Новгород: НГТУ, 2014 [54].

23) 2-й Международной научно-практической конференции "Современные тенденции и инновации в науке и производстве", г. Междуреченск, КузГТУ, 2014 [56].

24) 12-ой международной научно-технической конференции "Перспективные технологии в средствах передачи информации" – Владимир: ВЛГУ – 2017 [57,58].

Публикации. Основные положения диссертации опубликованы в 56 печатных работах, включая 12 публикаций в рецензируемых научных изданиях из перечня Минобрнауки РФ, издано учебное пособие (в соавторстве), получено 15 патентов на изобретение, 1 патент на полезную модель.

Реализация результатов работы. Использование результатов работы диссертационного исследования подтверждено актами об их внедрении полученными

от в/ч 32152, ФГУП НИИ "КВАНТ", в/ч №56707, НТЦ "Фобос НТ", ООО "Славсервис-Связь; использованы в учебном процессе Академии ФСО России г. Орел, Рязанского государственного радиотехнического университета, г. Рязань.

Личный вклад автора в публикациях с соавторами кратко характеризуется следующим образом: в публикациях [4, 9, 18, 37, 44, 57, 133, 137] представлен подход к построению моделей на основе выделения переменных длительностей сегментов анализа на однородных участках РС; в [6, 28, 45, 56] сформулированы технические решения по классификации сегментов РС, в [5, 7, 17, 20 – 25, 53, 128] разработаны теоретические принципы применения технологий построения систем обработки РС с переменной структурой и показаны прикладные возможности их реализации, в [10, 14, 15, 29, 31, 35, 38, 41- 43, 47, 48, 51, 52, 58, 131, 139, 143] раскрыты и представлены теоретические основы и практические особенности использования моделей и методов обработки РС на основе учета зависимости элементов декомпозиции РС при ЛП, в [11, 16, 33, 39, 50, 54, 141, 142] разработаны теоретические и практические аспекты полиспектрального анализа в приложении к задаче шумоподавления в РС, в [19, 34, 49, 135, 140] раскрыты особенности реализации процедуры векторного квантования (ВК) параметров ЛП РС, в [36, 40, 55] рассмотрены особенности расчета параметров передаточной функции голосового тракта при использовании ЛП, в [8, 26, 27, 30, 32, 38, 46, 129, 130, 132, 134, 136, 138] представлены практические аспекты реализации систем обработки РС в системах связи, в [12, 13] предложены подходы к проведению дальнейших исследований по построению индивидуально ориентированных систем обработки РС.

Структура и объем работы. Диссертационная работа изложена на 332 страницах текста, из них 292 страницы основного текста, содержит 79 иллюстраций и 21 таблицу, состоит из введения, шести разделов, заключения, списка сокращений, списка литературы (241 наименование), 2 приложений.

1. СУЩЕСТВУЮЩИЕ ПРОБЛЕМЫ АНАЛИЗА И ОБРАБОТКИ РЕЧЕВОГО СИГНАЛА В СИСТЕМАХ СВЯЗИ

1.1 Вводные замечания

Разработка и совершенствование устройств обработки РС функционирующих в системах связи является актуальной задачей, так как речь является естественной основой коммуникативного взаимодействия.

В данном разделе диссертации проанализированы существующие проблемы анализа и обработки РС в системах связи, выявлены факторы, определяющие актуальность задач обработки РС (подраздел 1.2), сформулированы основные требования, предъявляемые к системам обработки РС (подраздел 1.3), освещено состояние и структурированы проблемы обработки РС при его низкоскоростном кодировании в системах связи (подраздел 1.4), на основе проведенного анализа определены задачи диссертационного исследования (подраздел 1.5).

1.2 Анализ методов обработки речевого сигнала для задач связи

В настоящее время необходимо обеспечить потребителей современными телекоммуникационными и информационными услугами при сохранении соответствующей информационной безопасности [63, 73, 74]. Одной из таких услуг является предоставление качественной, конфиденциальной телефонной связи, которая немыслима без перехода к цифровой обработке и передаче РС.

В структуре инфокоммуникационной системы связи можно выделить следующие функционально взаимосвязанные между собой подсистемы: информационную, телекоммуникационную, а также систему обеспечения информационной безопасности.

Телекоммуникационную основу составляют существующие сети и комплексы связи, эволюционное развитие которых позволяет в настоящее время создать современную телекоммуникационную среду, отвечающую международным требованиям, обеспечивающую эффективность, безопасность и устойчивость обмена данными. Тенденции перехода к цифровым сетям связи и пакетной передаче данных позволили предоставить пользователям более широкий спектр инфокоммуникационных услуг. При этом речь до сих пор остается одним из основных средств общения, что подтверждает востребованность методов эффективной цифровой обработки РС [73, 149, 151].

Существующие цифровые тракты используются одновременно для передачи различных видов электросвязи (телефонии, передачи данных и др.) [86,151]. Практически это означает, что с введением единых протоколов взаимодействия сетей и систем, а также единой процедуры установления и разрушения соединений ликвидируются самостоятельные службы электросвязи для передачи (приема) РС, текста, данных и изображений. В такой сети важное внимание уделяется разработке методов эффективной обработки, кодирования данных, РС и видео для их компактного цифрового представления и дальнейшей передачи по КС [108, 112, 118, 158, 180]. С учетом постоянно возрастающих требований к устройствам обработки РС (повышению качества и снижению скорости передачи, уменьшению массогабаритных и стоимостных показателей, технологичности и унификации аппаратуры) целесообразно продолжение работ по улучшению данного класса устройств [198]. Это подтверждается и перспективами развития сети связи России. На рисунке 1.1 показана мультисервисная мультипротокольная сеть, основанная на принципе конвергенции всех видов связи [123, 126].

Основными свойствами данной сети являются:

– мультисервисность – возможность предоставления абоненту заданного перечня услуг связи (телефонная и видеотелефонная связь, передача данных, доступ к информационным ресурсам) в рамках одной системы связи;

- масштабируемость – возможность добавлять и исключать сервисы, наращивать абонентскую емкость и вычислительную серверную мощность без структурных изменений ядра системы;
- конвергентность – возможность сопряжения с существующими системами связи и информационного обеспечения различного назначения;
- безопасность – возможность обработки мультимедийной информации, содержащей сведения ограниченного распространения.

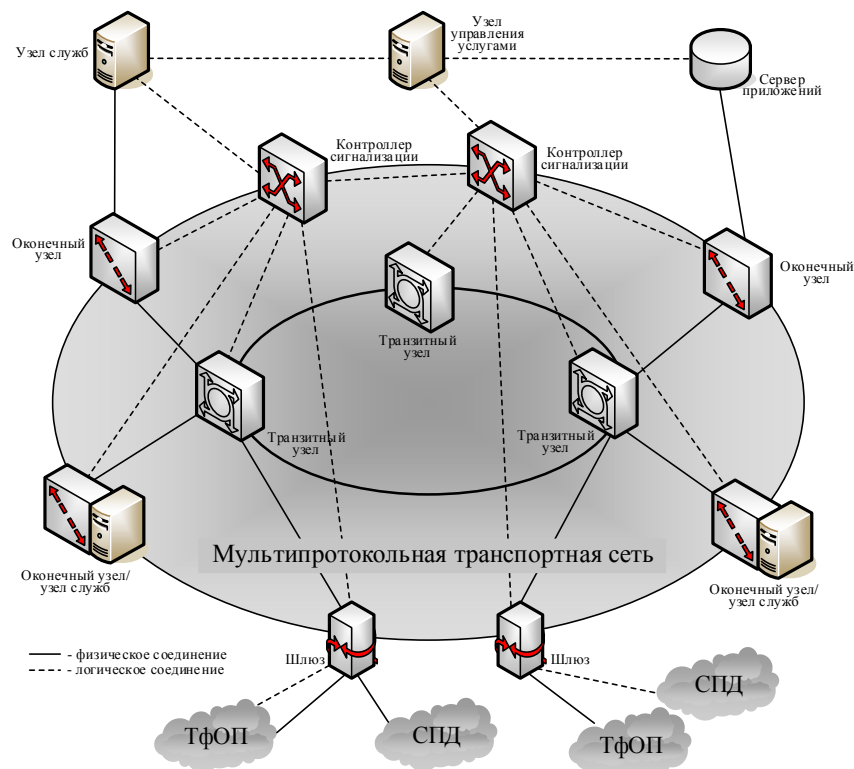


Рисунок 1.1 – Архитектура мультисервисной мультипротокольной инфокоммуникационной сети связи

Сопряжение с различными системами связи осуществляется в целях формирования общего информационно-коммуникационного пространства, являющегося технологической основой системы обмена информацией.

Таким образом, современная сеть связи основана на цифровых методах приема и передачи информации и коммутации пакетов данных. Современные технологии, на которых построена цифровая магистральная транспортная сеть, позволяют качественно и без потерь передавать РС практически на любой скорости. Слабым

местом такого построения, с точки зрения качественной обработки РС, являются линии привязки, ограничивающие пропускную способность КС в различных условиях функционирования. Анализ ситуаций предоставления инфокоммуникационных услуг по обработке РС выявил факт воздействия акустических шумов, что указывает на необходимость совершенствования устройств обработки речевой информации в части, касающейся их защиты от акустических воздействий. Условия функционирования систем обработки РС указывают на факты возможного зашумления в условиях акустических воздействий (УИАВ) с отношением сигнал-шум (ОСШ) на входе кодера менее 15 дБ, как природного (шум ветра, дождя), так и техногенного характера (шум двигателя автомобиля, вертолета, шум боя). Качественно новым этапом развития мобильных персональных абонентских терминалов (МПАТ) является введение средств обработки РС в условиях зашумления (рисунки 1.2) [61, 62, 75, 184].

Основу данного этапа составляет внедрение цифровых процессоров обработки сигналов (ЦПОС) и универсальных микропроцессоров с переходом на принцип обработки РС непосредственно в терминале абонента [103, 153, 172]. В современных системах связи с множественным доступом, основанных на одновременном совместном использовании сетевых ресурсов всеми активными абонентами, применение современных устройств обработки речевого сигнала (УОРС) с возможностью шумоподавления на входе кодера позволит в значительной степени повысить ОСШ при учете требований по субъективной оценке, снизив при этом требования к пропускной способности КС.

С одной стороны, современные системы обработки РС обладают потенциальными возможностями адаптации системы обработки РС к характеру и уровню акустического воздействия с целью шумоподавления, а с другой стороны, существует необходимость эффективного использования вычислительной мощности ЦПОС или универсальных микропроцессоров, применяемых в УОРС.

Наиболее важными характеристиками УОРС, применяемыми в МПАТ и влияющими на основные показатели, являются [188]:

- качество звучания синтезированной речи;
- скорость передачи речевой информации;

- алгоритмическая задержка РС в процессе его обработки;
- устойчивость к ошибкам в цифровом КС.

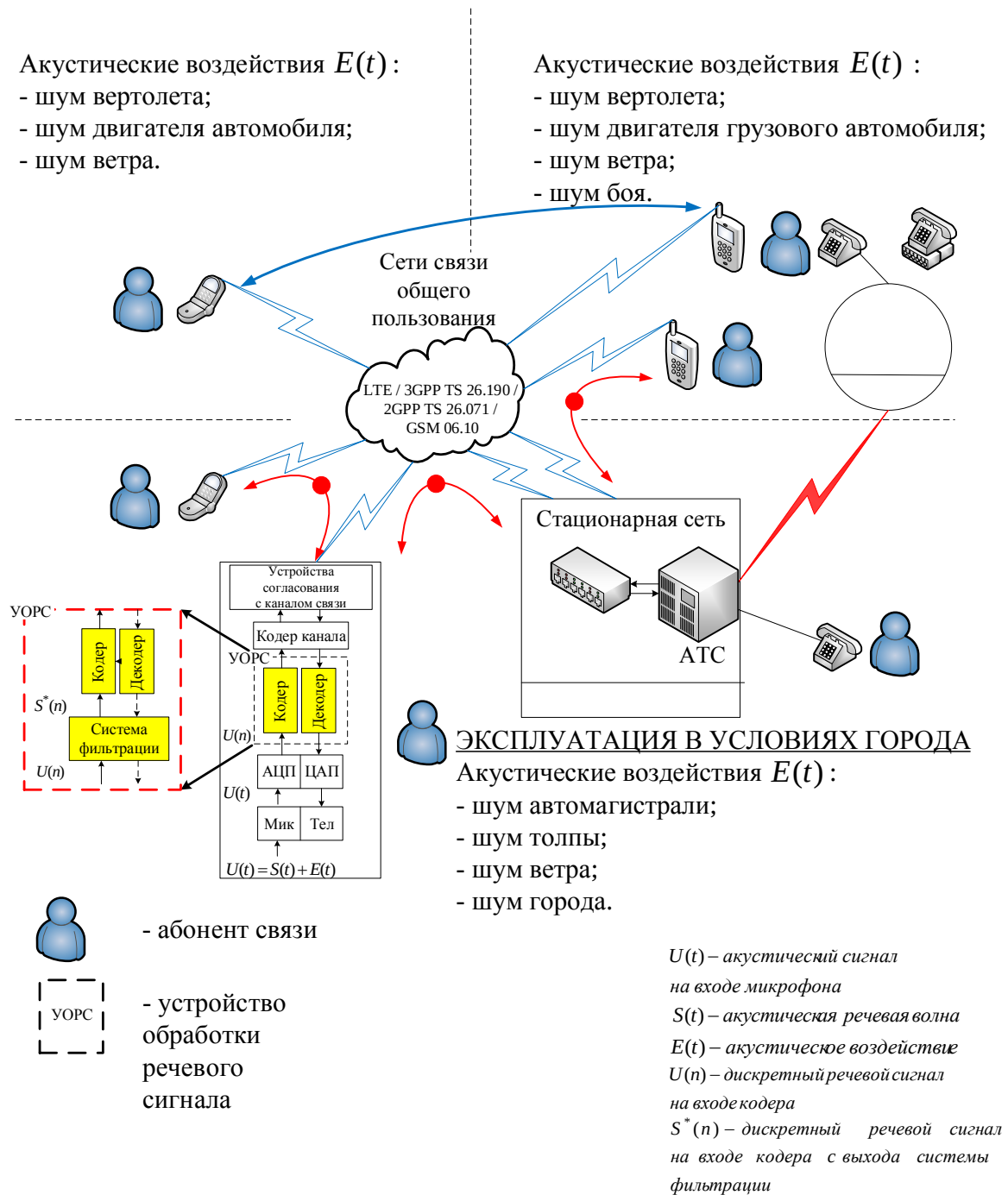


Рисунок 1.2 – Вариант фрагмента сети связи с использованием МПАТ

Качество речи, воспринимаемой абонентом, может быть оценено в соответствии со следующими показателями:

- разборчивость речи;

- естественность ее звучания и узнаваемость говорящего;
- ОСШ;
- интегральное качество речи.

Одним из элементов систем связи являются абонентские комплексы оперативной коротковолновой (КВ) радиосвязи, предназначенные для передачи данных [86, 158]. Повышение требований к оперативности и мобильности процессов управления приводит к необходимости пересмотра принципов построения и функционирования средств и комплексов КВ-радиосвязи. Принципиальной особенностью данного вида связи является необходимость предварительного низкоскоростного кодирования РС. При обеспечении связи с малыми подвижными объектами в условиях чрезвычайных ситуаций и вооруженных конфликтов телекоммуникационные услуги предоставляются широкому спектру должностных лиц.

Поскольку объем речевого трафика в системах инфокоммуникаций достаточно велик, то обработка и передача речевой информации невозможна без преобразования РС в компактную цифровую форму, выполняемого РПУ, что подтверждает актуальность работы по выбранному направлению исследования [219, 236]. Тактико-технические требования, предъявляемые к качеству функционирования РПУ в сетях связи, определяются требованиями по качеству обслуживания абонентов инфокоммуникационной системы [69, 79, 99].

Исследование и разработка перспективных методов обработки РС не может быть произведена без учета тенденций развития телекоммуникационной системы в мире, так как характеристики и свойства последней определяют требования к применяемым в ней системам аналого-цифрового (АЦП) и цифро-аналогового преобразования (ЦАП) РС.

Уровень требований, предъявляемых к системам и сетям связи, постоянно повышается по трем основным причинам:

- увеличение числа пользователей;
- расширение числа видов услуг, в которых заинтересованы пользователи;
- повышение уровня требований к качеству обслуживания (достоверности получаемой информации, времени ее доставки и безопасности).

В основе современных требований к системам телекоммуникаций лежит принцип интеграции, подразумевающий объединение систем электросвязи, их компонентов, технологий, видов, методов доставки и обработки различной информации. Основные направления развития процессов интеграции систем связи представлены ниже (рисунок 1.3). Данные направления определяют тенденции развития и совершенствования устройств обработки РС.

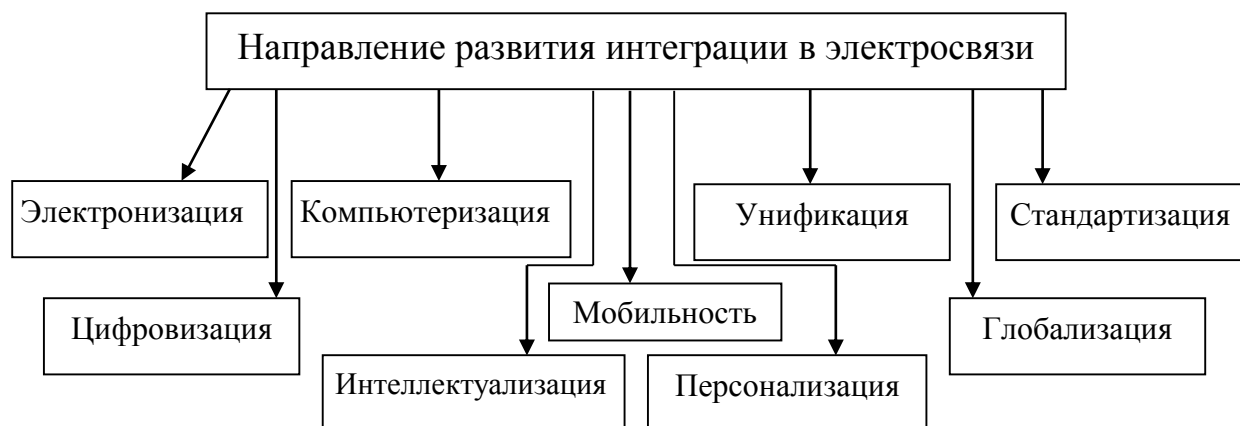


Рисунок 1.3 – Направления развития процессов интеграции в электросвязи

Появляются запросы на новые виды услуг, требующие для их осуществления широкополосного цифрового канала (ЦК). Это, прежде всего, видеосвязь, видеоконференция. При этом большинство из требуемых услуг являются услугами с комплексным представлением информации, которые в рекомендациях Международного союза электросвязи (МСЭ) определяются как мультимедиа.

Для эффективного информационного обмена абонентам необходимо, кроме качественной телефонной связи в разной акустической обстановке [3], предоставлять возможность передавать и принимать данные, видеоинформацию, обмениваться файлами в реальном масштабе времени, передавать сообщения по электронной почте, предоставлять возможность доступа к информационным базам и банкам данных, организовывать видеоконференции и др.

В настоящее время тенденцией развития систем связи является широкое использование информационных технологий в телекоммуникационных сетях. Пере-

ход к цифровым сетям и пакетной передаче данных позволил предоставить пользователям более широкий спектр защищенных услуг. При этом достаточно большую часть телетрафика в различных приложениях составляет передача РС, что также подтверждает актуальность работ по данному направлению. В перспективах развития взаимоувязанной сети связи Российской Федерации указывается на то, что в качестве базовой сети связи будет использоваться мультипротокольная защищенная сеть на основе IP (Internet Protocol), основанная на принципе взаимопроникновения (конвергенции) всех видов связи.

В основе современных сетевых технологий, обеспечивающих высокую эффективность использования полосы пропускания КС и минимальное значение задержки трафика в сети, лежит принцип коммутации пакетов и быстрого пакетного мультиплексирования. Такой подход позволяет избавиться от "простоев" в использовании канала [126]. Принципы пакетной передачи данных полностью соответствуют специфике возникновения и существования РС. Таким образом, исследования по разработке алгоритмов обработки РС для его низкоскоростного кодирования при переменной скорости передачи являются актуальными [201, 226, 227].

Сейчас доминирующим при создании и развитии инфокоммуникационных систем является асинхронный ввод информации, то есть пульсирующий трафик, основанный на принципе коммутации пакетов. Основным достоинством такого принципа является высокая общая пропускная способность сети при передаче пульсирующего трафика, то есть пропускная способность физических КС между абонентами перераспределяется динамически в соответствии с реальными потребностями их трафика, что весьма актуально при передаче различной мультимедийной информации.

Организация режима переменной скорости передачи при многоабонентской речевой обработке показана ниже (рисунок 1.4).

При передаче РС по сети возникают задержки, изменяющиеся случайным образом, что связано с изменением времени прохождения пакета через сеть, поэтому высока вероятность разделения или пропадания кадра передачи кодера. Проблема джиттера весьма существенна в сетях с пакетной коммутацией.

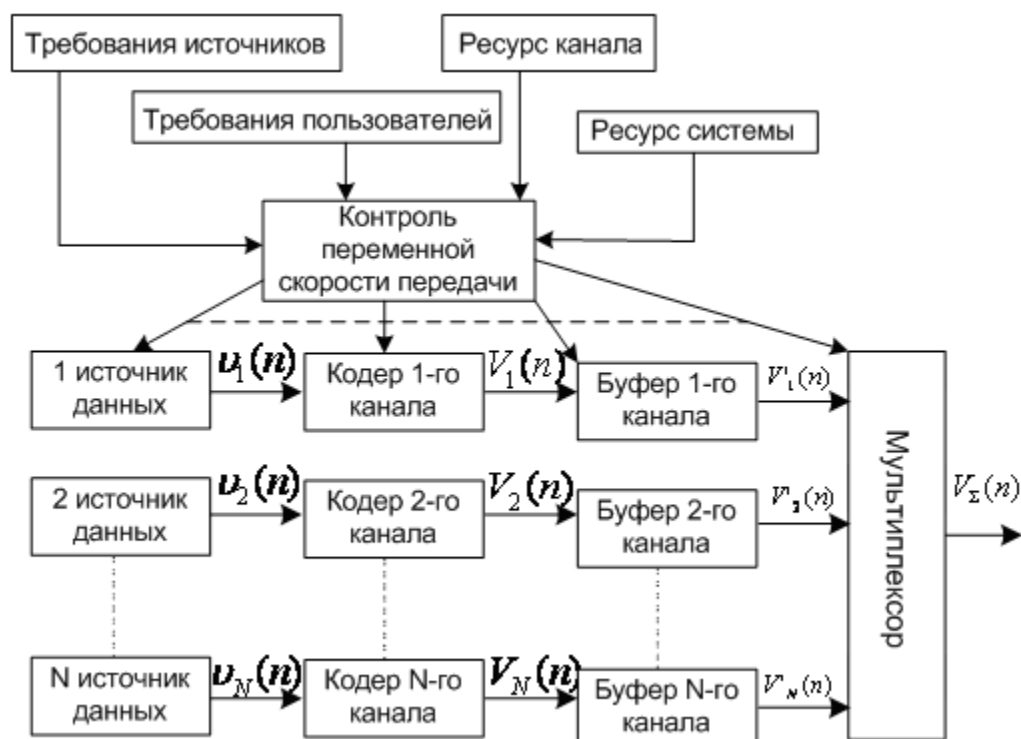


Рисунок 1.4 – Организация режима переменной скорости передачи

Одним из возможных решений, частично устраняющих данные недостатки, является использование технологии организации виртуальных каналов (VPN-соединений) (*Virtual Private Network*) совместно с технологией *MPLS* (*MultiProtocol Label Switching*), которая в последнее время получила широкое распространение. На основе протокола IP-MPLS строятся совершенно новые сети –NGN (*Next Generation Network*), способные обеспечить предоставление широкого спектра мультимедийных услуг [126].

Для предоставления пользователям сети связи "гибридных" информационно-телекоммуникационных услуг создаются универсальные многофункциональные терминальные устройства. Подобные многофункциональные устройства, как правило, имеют несколько режимов связи. В таких устройствах одной из актуальных задач является разработка качественных алгоритмов обработки РС. Применение подобных алгоритмов позволит упростить процедуру наращивания абонентского терминала для предоставления абоненту новых услуг и уменьшит его стоимость. МПАТ должны обеспечивать при этом высокое качество передачи речи и других видов информации, что даст возможность пользователю не быть "привязанным" к

месту постоянного расположения терминала связи. Это позволит выполнять задачи, стоящие перед системой связи в целом, и предоставлять абонентам расширенный перечень современных информационно-телекоммуникационных услуг, в том числе и по обработке РС [64]. Основными направлениями развития современной системы связи, определяющими тенденции совершенствования УОРС, являются [62, 82, 86, 149, 151, 155, 167, 168, 180]:

- полная цифровизация первичной (базовой) и вторичных (терминальных) сетей связи;
- активное развитие сетей мобильной радиосвязи на базе сотовых структур и глобальных спутниковых систем;
- всеобщую цифровизацию процессов формирования, обработки, коммутации и передачи РС;
- формирование рынка мультимедийных услуг и интеллектуальных сетей массового обслуживания, тесно связанное с тенденцией интеграции терминального оборудования;
- разработка новых принципов, способов и методов (технологии) организации различных видов связи;
- внедрение современной цифровой коммутационной техники и расширение возможностей существующей;
- создание мобильных устройств современных сетей связи;
- использование многофункциональных абонентских терминалов;
- увеличение пропускной способности линий связи в цифровых системах передачи информации;
- широкое использование радиодоступа и волоконно-оптической технологии на абонентских линиях связи;
- развитие сетей радиосвязи с малыми подвижными объектами;
- создание сетей универсальной персональной и ведомственной связи;
- интеллектуализация сетей передачи сообщений.

В комплексах телекоммуникаций КРС устанавливают в аппаратуре различных средств связи, он является составной частью УОРС [27]. В исследованиях,

направленных на разработку качественных систем обработки РС, активно развиваются сети подвижной радиосвязи, широко применяются услуги, предоставляемые *IP*-телефонией [74,123, 126, 138, 169]. Таким образом, УОРС может стоять как в канальной аппаратуре, так и встраиваться в телефонный аппарат или устанавливаться в системах мобильной связи [159, 161].

После преобразования аналогового РС в цифровую форму к нему применимы различные виды обработки [203]. В частности, существует множество методов сжатия оцифрованного сигнала и шумоподавления в нем, которые снижают требования к КС, необходимые для передачи речевых сообщений [209]. Основной задачей реализации процедуры качественного шумоподавления в РС является правильная оценка характеристик шумового воздействия, которая может быть решена применением технологий полиспектрального анализа [11].

Низкоскоростное кодирование РС и шумоподавление в нем приводит также к снижению качества воспроизводимой речи, что обусловлено следующими причинами. Во-первых, реализация алгоритмов обработки вызывает задержки при передаче РС. Для простых алгоритмов эта задержка вполне допустима и составляет единицы миллисекунд. Однако, для сложных алгоритмов, обеспечивающих значительное сжатие, длительность задержки может возрастать до нескольких десятков миллисекунд. Во-вторых, алгоритмы, обеспечивающие значительную степень сжатия, не допускают полного качественного восстановления сжимаемого сигнала, что также снижает качество воспроизводимой речевой информации. При этом возникают потери и задержки речевых пакетов данных, которые связаны с применением разных технологий и способов их обработки [126].

Системы обработки РС занимают одно из ключевых мест среди устройств, используемых при построении сетей связи. Системами обработки РС можно оснастить компьютеры, что позволит организовать цифровую связь с рабочего места пользователя с любым абонентом. Такие системы также можно использовать для создания компактных архивов различных речевых текстов [90].

Представленные выше тенденции развития современных инфокоммуникаций подтверждают актуальность поиска нового подхода к развитию методов обработки РС в режиме реального времени.

1.3 Требования, предъявляемые к системам обработки речевого сигнала

На первых этапах перехода к цифровым методам передачи и обработки РС в основном использовались методы непосредственного кодирования огибающей, в основе которых лежит теорема В. А. Котельникова о представлении финитных сигналов, ограниченных по спектру, последовательностью дискретных отсчетов [99, 167, 187]. Применение данных методов позволяет получить приемлемое качество цифрового представления РС, однако скорость передачи речевых данных в пределах 16 – 64 кбит/с не позволяет эффективно использовать пропускную способность ЦК передачи в некоторых приложениях связи [179, 187, 213, 219].

Алгоритмы вокодерного кодирования, такие как речееlementные и некоторые параметрические (формантные, полосовые), необходимого качества восприятия речи на приеме не достигают [160, 177]. С появлением в 80-х годах высокоскоростных микропроцессоров гибридные спектрально-временные вокодеры нашли свое широкое применение в различных приложениях, требующих компактного цифрового представления РС [60, 97, 193].

Среди многообразия разработанных методов обработки РС, применяемых в системах связи, одним из наиболее эффективных является метод ЛП [116, 152, 230]. Системы компрессии РС на их основе называются липредерами [161]. Тенденцией развития современных систем обработки РС является широкое применение в алгоритмах компрессии РС макропроцедур ЛП [158, 185, 188].

Перед разработчиками встает серьезная проблема улучшения показателей систем обработки РС в условиях наложенных ограничений на цену, качество обработки и представления выходных данных, скорость передачи, ограничений по применению разрабатываемых устройств. Анализ данных систем показывает, что в их развитии можно выделить несколько тенденций:

- сохранение качества обработки и представления выходных данных при снижении скорости передачи;
- повышение качества на заданных скоростях;
- расширение сфер применения результатов обработки;
- широкое использование разновидностей ЛП;
- усложнение алгоритмов обработки РС за счет применения различных адаптивных алгоритмов [175].

При анализе перспективности различных методов АЦП и ЦАП РС и формировании комплекса требований к перспективным системам обработки необходимо учитывать основные тенденции развития телекоммуникационных систем в мире [171, 199]. Выполненный анализ тенденций развития систем телекоммуникаций позволяет сформировать основные требования к системам обработки РС [188]:

1. Высокое качество обработанного РС A (близкое к качеству РС на выходе кодека G.726);
2. При кодировании слоговая разборчивость более 90 %;
3. Субъективное качество речи на выходе систем обработки – более 3,9 (Mean Opinion Score (MOS), ГОСТ Р 51061–97);
4. Используемая скорость передачи РС – не выше 9,6 кбит/с;
5. Возможность работы в широком динамическом диапазоне входных воздействий РС и акустических шумов;
6. Возможность обработки неречевых и фоновых звуков;
7. Возможность получения несмещенных оценок параметров РС с целью их дальнейшего анализа;
8. Низкая требуемая емкость устройств памяти R ;
9. Простота сопряжения со смежными кодеками мультимедийных абонентских терминалов;
10. Возможность реализации на современной элементной базе и учет тенденций ее развития;
11. Обработка речевых данных с показателями качества не ниже требуемых;

12. Высокая устойчивость к канальным ошибкам с вероятностью 10^{-3} (в соответствии с рекомендацией МСЭ SG15/WP2);
13. Возможность изменения скорости передачи;
14. Незначительное снижение качества речи при каскадном соединении систем обработки (не хуже, чем для кодека МСЭ G.726);
15. Низкая вычислительная сложность Q алгоритмов обработки;
16. Малая алгоритмическая задержка $\tau \leq 30$ мс РС в процессе обработки;
17. Малое энергопотребление и низкие массогабаритные показатели;
18. Низкая стоимость.

Традиционно задачи создания систем обработки РС с целью его кодирования и передачи в КС рассматриваются как оптимизационные в двух вариантах.

В первом задача обработки заключается в построении системы, минимизирующей скорость передачи V_m при сохранении заданного качества РС A .

Во втором случае ставится задача построения системы, обеспечивающей максимально высокое качество передачи РС A (минимум искажений) при заданной скорости ее передачи V_0 .

Дополнительно вводятся ограничения на стоимость системы, массо-габаритные показатели, вычислительную сложность Q , устойчивость к сбоям в КС A_n , величину задержки сигнала в системе τ и т. д.

- | | |
|--|-------------------------------------|
| 1. $V_m \rightarrow \min, V_m = \text{var};$ | 2. $A \rightarrow \max;$ |
| $A \geq A_0;$ | $A \geq A_0;$ |
| $\tau \leq \tau_0;$ | $V_m \approx V_0; V_m = \text{var}$ |
| $A_n \geq A_{0n}(P_C);$ | $\tau \leq \tau_0;$ |
| $Q \leq Q_0;$ | $A_n \geq A_{0n}(P_C);$ |
| | $Q \leq Q_0.$ |

Целевая функция представлена выражением

$$V_m |_{A \geq A_0; \vec{A}_a} \rightarrow \min.$$

Пространство $\vec{A}_a$ описывает ресурсы, которые могут использоваться для процедуры адаптации КРС.

1.4 Структуризация проблем обработки речевого сигнала

В системах связи наблюдается рост объема передачи данных, но при этом качественная и низкоскоростная передача РС в различных приложениях остается важной задачей. Актуальным также остается вопрос разработки речевого кодека, ориентированного на индивидуальные особенности речи пользователя и работающего в условиях акустических зашумлений [38].

Большинство вокодеров, используемых в системах связи, ориентированы на постоянную скорость передачи и основаны на методе ЛП [116]. В основе данного метода лежит предположение о локально-стационарной модели РС (рисунок 1.5) [100, 101, 102]. Это объясняется значительным качественным превосходством АЦП речи на основе метода ЛП над другими методами в диапазоне скоростей передачи менее 16 кбит/с [200].

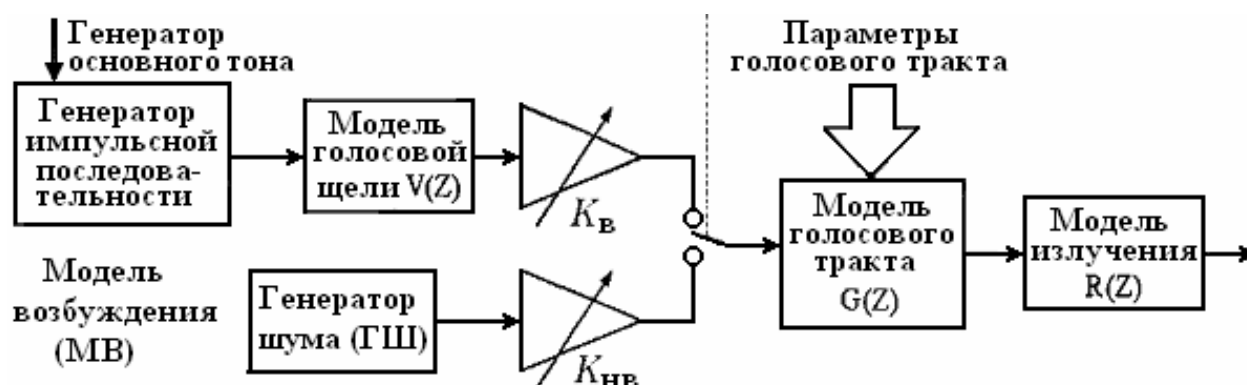


Рисунок 1.5 – Локально-стационарная модель формирования речевого сигнала

Отличительным признаком метода является одношаговое приближенное представление (предсказание) текущего отсчета $S(n)$ РС линейной комбинацией конечного числа предыдущих отсчетов.

Сущность метода ЛП заключается в том, что выборка РС $S(n)$ может быть предсказана линейной комбинацией предшествующих отсчетов этого сигнала:

$$S'(n) = \sum_{i=1}^M a_i S_{n-i} + e(n), \quad (1.1)$$

где $S'(n)$ – предсказанное значение РС; a_i – весовые коэффициенты линейного предсказания (КЛП); M – число коэффициентов или порядок ЛП.

Возникающая при этом ошибка предсказания находится по линейно-разностному уравнению (1.2), которое описывает функционирование фильтра анализа модели ЛП:

$$e(n) = S(n) - S'(n) = S(n) - \sum_{i=1}^M a_i S(n-i). \quad (1.2)$$

Задача анализа РС методом ЛП заключается в его фильтрации линейной системой с передаточной функцией вида:

$$A(z) = 1 - \sum_{i=1}^M a_i z^{-i}. \quad (1.3)$$

Обратная ей передаточная функция представляет собой фильтр синтеза и определяется соотношением:

$$H(Z) = \frac{1}{A(Z)} = \frac{1}{1 - \sum_{i=1}^M a_i Z^{-i}}. \quad (1.4)$$

Параметры РС изменяются с течением времени достаточно медленно, что позволяет рассматривать РС как стационарный на временных интервалах порядка 2,5–60 мс, называемых "окнами" или сегментами, при этом параметры передаточной функции голосового тракта также считаются неизменными, что дает возможность ввести в использование понятие "однородного сегмента РС". Таким образом, однородность сегмента РС связана с природой речеобразования и заключается в малом отличии параметров передаточной функции голосового тракта и сигнала

возбуждения при моделировании РС на основе метода ЛП. Возможность линейного прогнозирования текущего отсчета объясняется наличием значительных корреляционных связей между отсчетами РС при его равномерной дискретизации. Анализ выражений (1.1) – (1.4) говорит о том, что совпадение синтезированного РС с оригиналом обеспечивается при условии равенства множеств $\{a_i\}$ на этапах анализа и формирования РС и равенства порождающего сигнала и сигнала ошибки предсказания $e(n)$ в (1.2). Следует отметить, что последовательность, удовлетворяющая уравнению (1.2), является авторегрессионной, так как $S(n)$ совпадает с регрессией $S(n)$ по $S(n-1), S(n-2), \dots, S(n-M)$ или с математическим ожиданием $M\{S(n)/S(n-1), S(n-2), \dots, S(n-M)\}$.

Локальная стационарность ЛДДС на длительности сегмента анализа приводит задачу оценки параметров a_i по наблюдаемой реализации $S(n)$ к классу задач параметрической идентификации авторегрессионных стационарных систем во временной области [67, 80, 104, 105]. Выбор алгоритма идентификации в большой степени зависит от типа настраиваемой модели и выбора критерия, характеризующего качество идентификации [144, 182]. В липредерах в качестве настраиваемой модели, соответствующей (1.1) и (1.3), используется нерекурсивный цифровой фильтр M -го порядка [150]. В системах обработки РС, основанных на ЛП, порядок предсказания обычно является постоянным, а значения коэффициентов a_i фиксируются на коротких, примыкающих друг к другу и равных по длительности временных интервалах. При этом исходный РС рассматривается как совокупность сегментов стационарных случайных последовательностей, каждый из которых порождается системой с постоянными параметрами.

Специальные вычислительные приемы позволяют снизить нежелательное влияние переходных процессов, возникающих при переходе от одного сегмента к другому [117]. Такой подход основан на гипотезе о локально-стационарном характере процесса речеобразования, что позволяет оценивать изменения параметров или вероятностных характеристик нестационарного РС на основе моделей, инвариантных к временному сдвигу.

Следует заметить, что наибольшее влияние на качество предсказания имеют коэффициенты с меньшими номерами.

Одной из задач ЛП является идентификация коэффициентов синтезирующего фильтра. Критерием идентификации КЛП, как правило, выступает минимум средне-квадратической ошибки предсказания, выражающейся в виде:

$$M\{e^2(n)\} = M\left\{[S(n) - S'(n)]^2\right\}_{q_i} \rightarrow \min. \quad (1.5)$$

Тогда совокупность параметров a_i , минимизирующих (1.5) соответствует решению системы линейных уравнений:

$$\frac{\partial M\{e^2(n)\}}{\partial a_i} = 0, i = 1, 2, \dots, M, \quad (1.6)$$

которая выражается в аналитической форме через значения автокорреляционной функции $B(\tau)$, $\tau = 0, 1, 2, \dots, M$ исходного сигнала:

$$\begin{cases} B(0) \cdot a_1 + B(1) \cdot a_2 + B(2) \cdot a_3 + \dots + B(M-1) \cdot a_M = B(1) \\ B(1) \cdot a_1 + B(0) \cdot a_2 + B(1) \cdot a_3 + \dots + B(M-2) \cdot a_M = B(2) \\ B(2) \cdot a_1 + B(1) \cdot a_2 + B(0) \cdot a_3 + \dots + B(M-3) \cdot a_M = B(3) \\ \vdots \\ B(M-1) \cdot a_1 + B(M-2) \cdot a_2 + B(M-3) \cdot a_3 + \dots + B(0) \cdot a_M = B(M) \end{cases}, \quad (1.7)$$

Система уравнений (1.7) известна как система Юла-Уокера [1], и в матричной форме может быть записана в виде $[B] \times [a] = [R]$. Очевидно, что искомым вектор-столбец a будет определяться из выражения

$$[a] = [B]^{-1} \times [R], \quad (1.8)$$

где $\mathbf{B}^{-1}$ – матрица, обратная матрице $\mathbf{B}$.

$$\text{Матрица } \mathbf{B} = \begin{bmatrix} B(0) & B(1) & B(2) & \dots & B(M-1) \\ B(1) & B(0) & B(1) & \dots & B(M-1) \\ B(2) & B(1) & B(0) & \dots & B(M-3) \\ \dots & \dots & \dots & \dots & \dots \\ B(M-1) & B(M-2) & B(M-3) & \dots & B(0) \end{bmatrix}$$

является теплицевой [80], т. е. симметричной, в ней элементы главной диагонали и элементы, лежащие на параллелях к ней, соответственно равны между собой. Теплицева матрица является невырожденной (ее определитель отличен от нуля), поэтому система уравнений (1.7) является не только совместной, но и определенной, что означает единственность решения относительно $[a]$.

Известны разные способы ЛП, отличающиеся видом сигнала возбуждения и параметрами, описывающими состояния линейной формирующей системы [7, 186, 188]. Анализ различных подходов к выполнению базовых процедур данного метода позволил сформировать некоторые классификационные признаки, упорядочивающие его изучение (рисунок 1.6).

Метод ЛП РС использует модель в виде отклика линейной системы с переменными параметрами (голосового тракта) на соответствующий сигнал возбуждения (порождающий сигнал) [116]. Анализатор речепреобразующего устройства выделяет из короткого сегмента РС параметры состояния линейной системы и сигнала возбуждения, позволяющие синтезатору восстановить исходный сигнал с требуемой степенью верности, и эффективно представляет их в цифровой форме [217]. При этом совместно решаются задачи сжатия РС и его аналого-цифрового преобразования. В технике кодирования РС часто используются алгоритмы кодирования на основе модификаций метода ЛП, которые заключаются в различном представлении сигнала возбуждения [35, 186, 188, 231]. Основная задача, решаемая разработчиками, – это качественное и компактное представление параметров, характеризующих передаточную характеристику речевого тракта, а также сигнала возбуждения, аппроксимирующего ошибку ЛП [131]. В современных КРС широко применяется макропроцедура векторного квантования (ВК) [114, 221, 222, 224].

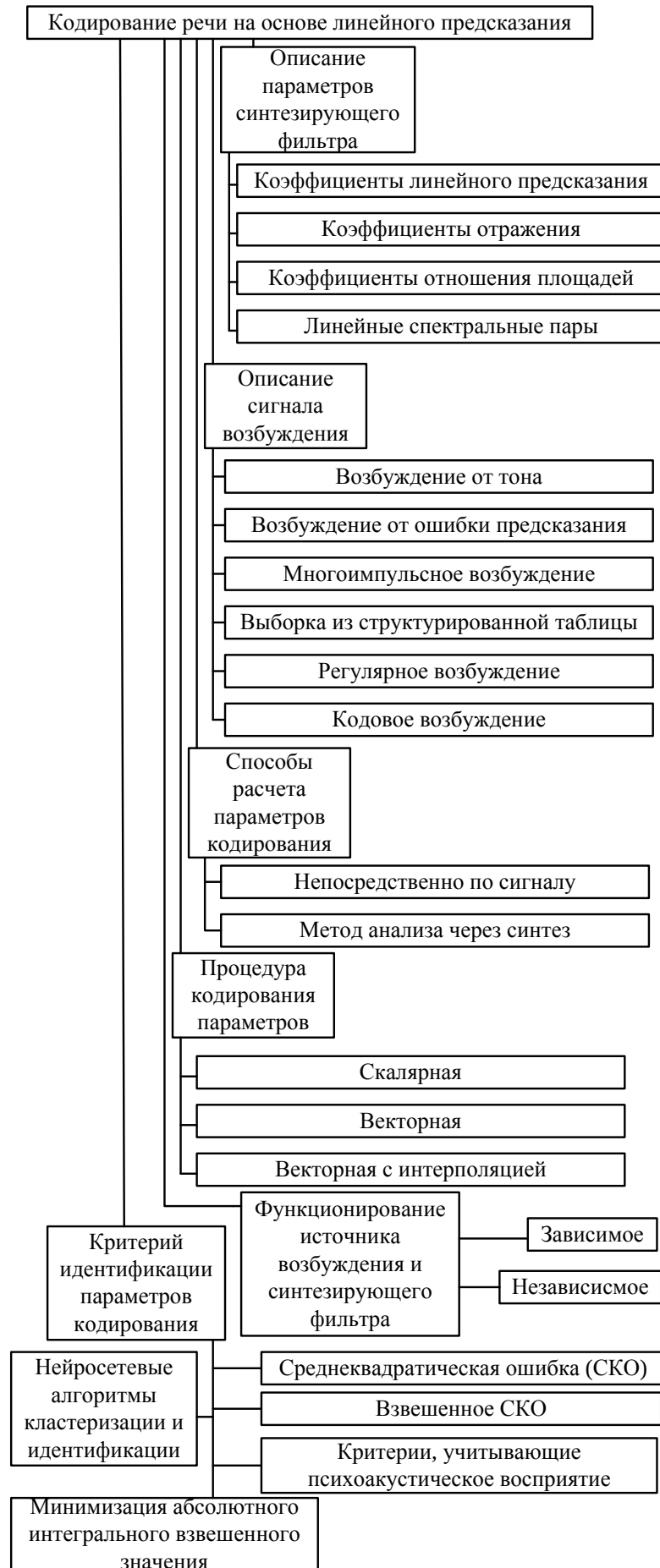


Рисунок 1.6 – Базовые процедуры линейного предсказания

Практика стандартизации и применения алгоритмов кодирования на скоростях $V \leq 16$ кбит/с, а также публикуемые результаты теоретических исследований в этой области убедительно показывают доминирующее положение методов ЛП и ВК при АЦП РС [158, 188, 195, 217].

Использование липредерных алгоритмов в различных телекоммуникационных приложениях показано в (таблице 1.1).

Таблица 1.1 – Основные стандарты кодирования речевого сигнала на основе метода линейного предсказания

Область применения	Стандарт				Скорость кодирования, кбит/с
	Наименование	Год	Тип липредера	Организация	
Телефонные сети общего пользования	G.728	1992	LD-CELP	Сектор стандартизации Международного Союза электросвязи	16
	G.729	1995	CS-ACELP		8
Сотовые системы радиосвязи	GSM full-rate	1989	RPE-LTP-LPC	Европейский институт стандартов в области связи	13
	GSM half-rate	1994	ACELP		6,5
	GSM-EFR	1995	ACELP		13
	IS-54	1991	VSELP	Ассоциация телекоммуникационной промышленности США	7,95
	IS-96	1996	CELP		2,4...9,6
	JDC full-rate	1991	VSELP	Министерство почты и связи Японии	6,7
	JDC half-rate	1993	PSI-CELP		3,45
Транкинговые системы радиосвязи	TETRA	1994	ACELP	Европейский институт стандартов в области связи	4,56
Спутниковые системы	Iridium		VSELP	Компании – разработчики оборудования	2,4
	Global star		CELP		1,2, 2,4,...,9,6
	Inmarsat-aero	1989	MPELP		9,6
Интернет-телефония	G.723.1	1995	A/MP-MLQ CELP	Сектор стандартизации Международного Союза электросвязи	5,27
Системы специального назначения, в т.ч. - военные	FS-1015	1984	LPC	Министерство обороны США	2,4
	FS-1016	1990	CELP		4,8

Типы представленных липредеров имеют в своей основе различные модификации метода ЛП:

1) с возбуждением от импульсов основного тона (ОТ), часто называемый просто *LPC (Linear Predictive Coding)*;

2) с многоимпульсным возбуждением – *MPELP (Multi Pulse Excited Linear Prediction)*;

- 3) с возбуждением от остатка – *RELP* (*Residual Excited Linear Prediction*);
- 4) с возбуждением от суммы векторов – *VSELP* (*Vector Sum Excited Linear Prediction*);
- 5) с возбуждением от кода – *CELP* (*Code Excited Linear Prediction*);
- 6) с регулярным импульсным возбуждением и долговременным ЛП – *RPE-LTP-LPC* (*Regular Pulse Excited Long Term Prediction Linear Prediction Coding*), применяемый в *Global System for Mobile Telecommunications* (GSM);
- 7) с возбуждением от алгебраического кода – *ACELP* (*Algebraic Code Excited Linear Prediction*);
- 8) с возбуждением от алгебраического кода и сопряженной структурой квантователя коэффициентов усиления сигналов возбуждения – *CS-ACELP* (*Conjugate-Structure and Algebraic Code Excited Linear Prediction*);
- 9) с возбуждением от кода и малой алгоритмической задержкой – *LD-CELP* (*Low Delay Code Excited Linear Prediction*);
- 10) с возбуждением от кода, отображающего сигнал возбуждения, синхронный с ОТ – *PSI-CELP* (*Pitch Synchronous Innovation*).

Первые реализации липредеров, такие как *LPC*-вокодер, способствовали передаче РС на низких скоростях – 2,4 и 4,8 кбит/с. На скорости 2,4 кбит/с обеспечивается приемлемый уровень разборчивости речи, но качественные характеристики, такие как естественность и узнаваемость синтезированной речи недостаточны. В кодерах с ЛП и возбуждением от остатка предсказания *RELP* сигнал ошибки (остатка) ограничивается по частоте и прореживается. Кодеры с многоимпульсным возбуждением (*MPE LPC* – *Multi Pulse Excitation Linear Prediction Coding*) используют вместо сигнала-остатка искусственную последовательность возбуждения РС на некотором временном интервале, параметры которой передаются в декодер.

Более сложные алгоритмы используют замкнутый *LPC*-кодер (называемый также кодером "анализа-через-синтез" – *Analysis-by-Synthesis* (*AbS*)), который характеризуется самооптимизирующейся процедурой поиска. Примеры стандартных замкнутых *LPC*-алгоритмов – это метод ЛП с кодовым возбуждением *CELP* [111, 190, 202, 206, 210, 211, 212, 218, 232], метод регулярного импульсного возбуждения

RPE LTP-LPC [188], используемый в европейских сотовых системах связи на скорости 13 кбит/с, и метод *LD-CELP* с низкой задержкой, принятый МСЭ в качестве стандарта кодирования РС *G.728* на 16 кбит/с [188]. Широкое распространение для различных приложений получило и множество других методов кодирования, в частности варианты адаптивного кодирования с предсказанием (*Adaptive Predictive Coding – APC*), разработанные в лабораториях компании *Bell*; метод ЛП с векторным возбуждением *VSELP*, предложенный фирмой *Motorola* в качестве стандарта для цифровых сотовых систем США, работающих на скорости 8 кбит/с; метод ЛП с предиктивным кодовым возбуждением (*Predictive Code-Excited Linear Prediction – PCELP*), созданный *DSP Group* в 1992г. [188]

Недостатками данных систем являются фиксированная длительность сегмента анализа РС, используемая при реализации данных технических решений, а также независимое представление и описание множеств элементов декомпозиции РС, вычисляемых с использованием модели ЛП в классе CELP подобных систем. Основными причинами применения вокодеров CELP в классе устройств кодирования РС в диапазоне скоростей $V \leq 16$ кбит/с являются:

1. Высокая степень согласованности анализа речи на основе метода ЛП с природой РС.
2. Локально-стационарный характер модели речеобразования, используемой в вокодерах с ЛП, достаточно хорошо соотносится с квазистационарным характером РС на временных интервалах, соответствующих режиму установившихся звуков.
3. Эффективное разделение передаточной функции голосового тракта и возбуждающей функции при анализе РС методом ЛП создает хорошие условия для сжатия РС с целью его качественного представления при низких скоростях передачи.
4. Важной особенностью передаточной функции голосового тракта является то, что она, в основном, характеризуется резонансами, которые, в свою очередь, достаточно хорошо представляются с помощью полюсной модели.

5. Высокая степень соответствия липредерных алгоритмов речепреобразования современному уровню развития теории и технологии ЦОС. Наличие эффективных расчетных алгоритмов реализации процедур анализа РС на основе метода ЛП обеспечивает достижение высокой точности анализа в режиме реального времени при соблюдении ограничений на вычислительную сложность, обусловленных современным уровнем развития цифровых процессоров обработки сигналов (ЦПОС).

6. Значительное качественное превосходство АЦП речи на основе метода ЛП с ВК параметров липредера над другими методами в диапазоне скоростей кодирования 1,2–16 кбит/с.

Следовательно, применение метода ЛП, использующего полюсную модель линейной дискретной динамической системы (ЛДДС), позволяет сохранить многие важнейшие характеристики передаточной функции речевого аппарата человека. При этом потери, получаемые при переходе от полюсно-нулевой модели к полюсной, вполне компенсируются существенным упрощением алгоритмов обработки РС [116].

Высокая степень востребованности во многих телекоммуникационных приложениях указанного скоростного диапазона сопровождается его качественной недостижимостью для речепреобразующих устройств, основанных на методах непосредственного кодирования РС, речееlementных и спектральных параметрических вокодеров. К сожалению, в настоящее время наблюдается замедление развития данного класса речепреобразующих устройств, что связано с достижением определенных качественных показателей, удовлетворяющих большинство потребителей подобных систем в данном диапазоне скоростей.

Актуальной во многих приложениях телекоммуникаций остается задача снижения скорости передачи РС при сохранении качественных показателей синтезированной речи [53]. При определении характеристик кодера в рамках УОРС МПАТ качество синтезированной речи принято оценивать разборчивостью. В (таблице 1.2) приведены оценки слоговой разборчивости основных стандартов КРС, используемых в системах инфокоммуникаций, данный параметр оценивается в соответствии с [83, 84].

Исследования в данной области [96] позволяют говорить о том, что существующие УОРС комплексов инфокоммуникаций в условиях отсутствия зашумления обеспечивают слоговую разборчивость речи на уровне 85–90 %. При такой и более высокой разборчивости речи переговоры ведутся без переспросов. Понимание речи не требует от абонента какого-либо напряжения. Однако при ОСШ на входе кодера менее 15 дБ наблюдается снижение разборчивости речи и субъективной оценки качества из-за наличия шумов в обрабатываемом РС, что требует дополнительного напряжения слухового аппарата человека с целью выделения полезного информационного сигнала [61, 177, 183, 184, 196, 225, 233].

Слоговую разборчивость измеряют по методике и таблицам слогов, приведенным в [83, 84]. Бригада операторов состоит из дикторов (не менее 3 человек — мужчин и женщин) и аудиторов (4—5 человек), при этом тренировка бригады операторов не производится. Цикл измерений состоит из передачи всеми дикторами по 10 таблиц каждым и приема всеми аудиторами всех переданных таблиц. Слог считают неправильно принятым, если он принят неправильно, пропущен или добавлен.

Таблица 1.2 – Слоговая разборчивость речи
в зависимости от вида алгоритма кодирования

Характеристики кодера		
Алгоритм кодирования	Слоговая разборчивость речи, %	Скорость передачи, кбит/с
ACELP	85	4.8
RPE-LTP-LPC	90	13
Adaptive Multi-Rate (AMR)	85	6.4
Adaptive Multi-Rate Wideband (AMR WB)	90	15

На основе представленного построения КРС можно сделать вывод о том, что в настоящее время актуальным направлением в области кодирования РС является разработка адаптивных алгоритмов, ориентированных на низкоскоростные каналы передачи. Основной особенностью адаптивных алгоритмов является обеспечение

изменения параметров и/или структуры объекта управления (в данном случае – кодека) в соответствии с результатами анализа входных данных [147]. Фиксированное структурирование пространства кодируемых параметров и постоянство мощностей подпространств представлений (порядок предсказания для ЛП), присущие существующим алгоритмам и выражающиеся в фиксированной структуре кодеков, не создают необходимых условий для максимального использования имеющейся априорной и апостериорной информации о РС, что препятствует более глубокой оптимизации кодека.

В связи с этим перспективным направлением при создании новых КРС является разработка адаптивных алгоритмов кодирования РС, обеспечивающих функционирование структурно и параметрически адаптивных систем обработки РС [94,113, 175, 192]. Основной задачей при работе такого кодека является выбор рациональной структуры кодера и его параметров на этапе анализа и передача параметров выбранной структуры на декодер. Представляется перспективным для создания адаптивного кодека с переменной структурой использовать результаты исследований по внедрению кодека на основе CELP модели, наиболее часто используемой при низкоскоростном кодировании на основе метода ЛП, так как в ней заложено большинство перспективных решений [202, 206, 217, 218]. В наиболее востребованном для решения указанных задач диапазоне скоростей (до 16 кбит/с) получили развитие гибридные методы кодирования на основе метода ЛП.

Недостатками представленных способов является относительно высокая скорость передачи при заданном качестве РС, а также значительный расход информационного ресурса на представление параметров, описывающих сигнал возбуждения формирующего фильтра голосового тракта. Это объясняется тем, что выделение и обработка этих параметров осуществляются на каждом фиксированном интервале квазистационарности РС. На данном участке сигнал принято считать близким к стационарному, вследствие чего он получил название квазистационарный. При ЛП по КС передается информация о коэффициентах формирующей модели и параметрах, характеризующих сигнал возбуждения и обрабатываемый РС, при

этом данные параметры кодируются независимо друг от друга. Данный факт негативно сказывается на возможности дальнейшей оптимизации.

На практике используют фиксированные сегменты РС, на основании которых рассчитывают параметры состояния линейной системы и сигнала возбуждения, позволяющие синтезатору восстановить исходный сигнал с требуемой степенью верности [214]. При этом входные отсчеты РС разделяются на сегменты фиксированной длины (80, 160 или 320 отсчетов, что при частоте дискретизации 8кГц соответствует 10, 20 и 40мс соответственно). Для каждого из сегментов рассчитывают коэффициенты формирующей модели, а затем обрабатывают с использованием алгоритма ЛП с возбуждением от кода, так как декомпозиция РС на локальных временных интервалах достаточно хорошо соотносится с квазистационарной моделью РС на временных интервалах, соответствующих режиму установившихся звуков. При этом происходит разделение передаточной функции голосового тракта и возбуждающего сигнала. Недостатком такого подхода является использование фиксированной длительности сегмента анализа, что не соотносится со спецификой возникновения и существования РС. В 80-х XX века Атал в своих работах попытался ввести фиксированные значения переменных длин сегментов анализа [189], однако в то время данный подход натолкнулся на сложности при осуществлении синхронизации. Переход к линейной полюсной модели речеобразования и формирование РС в виде реакции рекурсивной линейной дискретной динамической системы с сосредоточенными параметрами на коротких временных интервалах является недостатком ЛП РС, так как он не учитывает конечность во времени РС и его нестационарную природу. Также недостаточно полно используются существующие зависимости параметров декомпозиции РС, возникающие в процессе применения процедуры ЛП.

Для повышения качества синтезированного РС разработчики систем обработки на основе ЛП усложняют представление сигнала возбуждения, чтобы, с одной стороны, компактно передать его на приемную сторону, а с другой – приблизить его к виду ошибки предсказания, как идеальному сигналу воздействия на

фильтр синтеза [188, 194]. Дополнительно вводятся разные варианты квантования параметров липредера (скалярное, векторное, каскадное векторное) [114, 188].

Так как РС является нестационарным конечным во времени процессом, то более правильной его моделью является реакция системы с переменными во времени параметрами на участках, соизмеримых с частотой дискретизации сигнала. При этом переменный характер состояния системы можно нацелить на повышение эффективности представления РС за счет доступной степени использования нестационарных свойств речи и особенностей его нелинейного моделирования [178]. Однако такое представление, являясь по сути более правильным с точки зрения обработки РС, не находит применения на практике в связи со сложностью и ограниченностью применения таких подходов [112, 156, 170]. В (таблицах 1.3 и 1.4) представлены значения средней длительности гласных и согласных звуков (фонем) [69].

Таблица 1.3 – Средняя длительность гласных звуков

Звук	Длительность гласных звуков, с						Среднее значение длительности, с	
	В словах				Во фразах			
	В середине слова		В начале и конце слова		Произвольное положение			
	Ударные	Неударные	Ударные	Неударные	Ударные	Неударные	Ударные	Неударные
у	0,19	0,11	0,25	-	0,17	0,1	0,2	0,11
о	0,22	-	0,25	-	0,17	0,12	0,2	0,14
а	0,23	0,1	0,26	0,18	0,18	0,08	0,21	0,08
э	0,2	0,08	0,25	0,19	0,14	0,11	0,18	0,12
и	0,18	0,1	0,26	0,16	0,16	0,09	0,19	0,11
ы	0,19	0,1	-	0,17	0,18	0,08	0,19	0,11

Таблица 1.4 – Средняя длительность согласных звуков

Тип согласных звуков		Длительность согласных звуков, с					
		Твердые			Мягкие		
Смычковые	Звонкие	б	д	г	бь	дь	гь
	Глухие	0,10	0,10	0,07	0,09	0,09	0,10
Щелевые	Звонкие	п	т	к	пь	ть	кь
	Глухие	0,02	0,03	0,03	0,03	0,10	0,04
Щелевые	Звонкие	в	ж	з	вь	жь	зь
	Глухие	0,10	0,11	0,11	0,08	-	0,13
Сонаты	Носовые	ф	ш	с	х	фь	шь
	Щелевые	0,15	0,14	0,14	0,12	0,12	0,24
Дрожащие	Носовые	м	н		мь	нь	
	Щелевые	0,09	0,09		0,09	0,10	
Аффрикаты	Дрожащие	л			ль	ль	
	Дрожащие	0,08			0,08	0,08	
Аффрикаты	Дрожащие	р			рь		
	Дрожащие	0,06			0,04		
Аффрикаты	Дрожащие	ц			ч		
	Дрожащие	0,14			0,09		

Приведенные в (таблицах 1.3, 1.4) значения превосходят принятую фиксированную длительность сегмента анализа РС в системах обработки, это дает предпосылки для снижения информативной избыточности параметров РС, используемых при реализации алгоритмов низкоскоростного кодирования. РС по своей природе обладает статистической и психоакустической избыточностью (рисунок 1.7) [181, 220]. В современных системах телекоммуникаций для повышения натуральности восприятия речи используется широкополосный РС, при этом эффективная полоса частот обрабатываемого РС увеличивается до 7 кГц (рисунок 1.8), что не всегда оправдано с точки зрения эффективности функционирования систем связи, приводя к нерациональному использованию ресурса доступной пропускной способности КС, решения по устранению данного недостатка представлены в [8, 134].



Рисунок 1.7 – Классификация составляющих избыточности речи

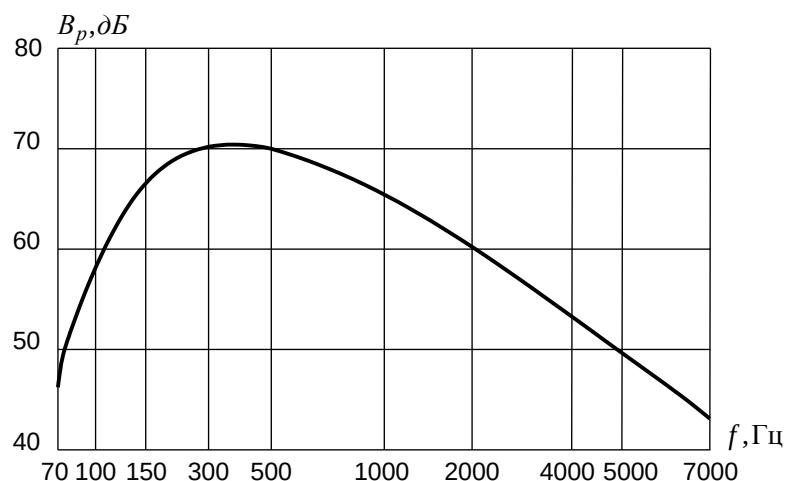


Рисунок 1.8 – Типовой спектр речевого сигнала

При использовании алгоритмов шумоподавления РС также применяется разделение на сегменты фиксированной длины с целью их дальнейшей обработки. Используемые при этом подходы, как правило, основаны на классическом представлении РС в спектральной области, что не позволяет эффективно разделять сегменты активной зашумленной речи и сегменты паузы, в которых присутствует шумовое воздействие [11, 54].

Таким образом, можно выделить основные недостатки существующих систем обработки РС:

- использование фиксированных сегментов анализа при реализации систем обработки РС, на которых предполагается наличие квазистационарных участков не соответствует особенностям возникновения и существования РС;
- переход к линейной полюсной модели речеобразования и формирование РС в виде реакции рекурсивной линейной дискретной динамической системы с сосредоточенными параметрами на коротких временных интервалах;
- расширение спектральной области обрабатываемого сигнала до 7 кГц без учета специфики образования отдельных фонем;
- использование сети передачи данных, не адаптированной к методам обработки и передачи РС;
- применение классического представления РС в спектральной области;
- отсутствие учета в перспективных системах обработки РС индивидуальных особенностей говорящего.

1.5 Постановка задач диссертационного исследования

Основываясь на вышеизложенных недостатках существующих систем обработки РС и учитывая существующие возможности ЦОС, экономические ограничения по стоимости УОРС и стремление к использованию известного математического аппарата, можно сформулировать основополагающее противоречие между существующими требованиями по улучшению характеристик функционирования систем обработки РС в условиях наличия акустических шумовых воздействий при ограничениях на скорость передачи и невозможностью их выполнения на основе

существующих технологий его обработки и методов анализа, не учитывающих особенностей формирования и моделирования РС в системах телекоммуникаций.

Для разрешения данного противоречия необходимо сформулировать основополагающую **гипотезу исследования**: использование положений теории линейной фильтрации и результатов полиспектрального анализа РС в условиях наличия акустических шумовых воздействий, а также учет зависимости элементов его декомпозиции и особенностей формирования РС при его моделировании в системах связи, дают возможность улучшить характеристики функционирования систем его обработки.

Структура и предполагаемые научные результаты исследования представлены на (рисунок 1.9).

ЭЛЕМЕНТЫ ТЕОРИИ АНАЛИЗА РЕЧЕВОГО СИГНАЛА ПРИ СТРУКТУРНО-ПАРАМЕТРИЧЕСКОЙ АДАПТАЦИИ СИСТЕМ ОБРАБОТКИ					
Понятийно-терминологический аппарат предметной области					
Аналитическая модель структурно-параметрической адаптации системы обработки речевого сигнала					
Методология структурно-параметрической адаптации систем обработки речевого сигнала					
Математический аппарат теории структурно-параметрической адаптации	Принципы и подходы к структурно-параметрической адаптации	Теоретическое обоснование возможности выделения однородных участков речевого сигнала одинаковой природы образования и вычисления параметров модели на них		Теоретические основы определяющие взаимозависимость элементов декомпозиции речевого сигнала при его кодировании на основе метода линейного предсказания при условии наличия акустических шумовых воздействий	
Математические модели систем обработки речевого сигнала при его низкоскоростном кодировании					
МЕТОДЫ ОБРАБОТКИ РЕЧЕВОГО СИГНАЛА, ОСНОВАННЫЕ НА МАТЕМАТИЧЕСКОМ АППАРАТЕ СТРУКТУРНО-ПАРАМЕТРИЧЕСКОЙ АДАПТАЦИИ					
Методы выделения однородных сегментов речевого сигнала одинаковой природы образования и вычисления параметров синтезирующей системы на них		Метод выделения взаимозависимых элементов декомпозиции речевого сигнала при его кодировании на основе метода линейного предсказания		Метод шумоподавления в речевом сигнале при наличии побочных акустических воздействий	
КОМПЛЕКС АЛГОРИТМОВ АНАЛИЗА И ОБРАБОТКИ РЕЧЕВОГО СИГНАЛА, ОСНОВАННЫЙ НА ИСПОЛЬЗОВАНИИ ДАННЫХ МЕТОДОВ					
Алгоритм определения однородных участков речевого сигнала одинаковой природы образования	Алгоритм адаптивной цифровой фильтрации РС в условиях интенсивных шумовых акустических воздействий	Алгоритм вычисления и коррекции параметров синтезирующей системы на сегментах одинаковой природы образования	Алгоритм поиска взаимозависимости элементов декомпозиции речевого сигнала при его кодировании на основе метода линейного предсказания	Алгоритм кодирования речевого сигнала при учете взаимозависимости элементов его декомпозиции при обработке на основе метода линейного предсказания	Алгоритм поиска активных участков речевого сигнала и его обработки в условиях интенсивных акустических зашумлений
АППАРАТНО-ПРОГРАММНЫЕ РЕШЕНИЯ ПО РЕАЛИЗАЦИИ РАЗРАБОТАННЫХ МЕТОДОВ И АЛГОРИТМОВ ОБРАБОТКИ РЕЧЕВОГО СИГНАЛА					

Рисунок 1.9 – Структура и научные результаты исследования

Таким образом, необходимо развить теоретические принципы анализа и моделирования РС при наличии акустических шумов, разработать методы и алгоритмы его анализа и обработки, учитывающие особенности формирования РС и возможности современных систем телекоммуникаций.

С учетом вышеизложенного, **частные задачи** диссертационной работы формулируются следующим образом:

- проанализировать актуальные пути развития систем обработки РС, учитывающие современные требования к телекоммуникациям;
- развить теоретические принципы анализа РС;
- разработать математические модели, методы и алгоритмы его анализа и обработки, в том числе при наличии акустических шумов, учитывающие особенности формирования РС и принятые технологии его обработки и позволяющие осуществить снижение средней скорости передачи данных в КС;
- разработать методологию снижения средней скорости передачи данных;
- разработать рекомендации по технической реализации и применению новых методов обработки РС в системах инфокоммуникаций.

1.6 Выводы по первому разделу

1 При учете современных тенденций развития телекоммуникаций необходимо комплексное техническое решение по формированию методологии понижения скорости кодирования РС при его передаче с сохранением качественных показателей синтезированной речи в условиях наличия акустических шумовых воздействий.

2 Достаточно часто при ведении переговоров в системах связи осуществляется обработка зашумленного РС, поэтому актуальным направлением является решение задачи шумоподавления в системах передачи и обработки РС.

3 КРС во многом определяют полноту услуг, предоставляемых абонентам, и характеризуют особенности функционирования систем инфокоммуникаций, задавая верхнюю потенциальную границу качества телефонной связи.

4 Важными типами перспективных кодеков, используемых в настоящее время в системах инфокоммуникаций, являются системы низкоскоростной передачи РС (1,2 – 9,6 кбит/с) на основе модификаций метода ЛП с ВК параметров липредера.

5 Целесообразно разработать систему обработки РС путем совершенствования решений, в том числе процедуры анализа через синтез, положенных в основу КРС на основе CELP модели.

6 Рассматриваемые системы обработки РС являются параметрически адаптивными, что с учетом нестационарности РС дает основание для их совершенствования на основе перехода к структурно-параметрической адаптации.

7 Локально-стационарный характер представления РС в существующих системах обработки на основе ЛП предполагает использование фиксированных длительностей анализируемых сегментов РС. Совершенствование процедур анализа РС, учитывающее особенности природы речеобразования, позволит перейти к выделению сегментов анализа переменной длительности, что даст возможность более адекватно производить моделирование РС и уменьшить количество бит, выделяемых для представления параметров передаточной функции голосового тракта при его кодировании на основе метода ЛП.

8 Системы обработки РС на основе ЛП с ВК, применяемые в современных инфокоммуникационных системах, рассматривают кодируемые параметры РС как независимые. Дальнейшее совершенствование этого класса систем может быть достигнуто на основе перехода к учету существующих зависимостей между указанными параметрами, что адекватно природе речеобразования и особенностям его моделирования в липредерах. Такой подход создаст условия для более рационального распределения количества бит, выделяемых для представления кодируемых параметров.

2 РАЗРАБОТКА МЕТОДА ВЫДЕЛЕНИЯ СЕГМЕНТОВ АНАЛИЗА ПРИ МОДЕЛИРОВАНИИ ПРОЦЕССОВ ОБРАБОТКИ РЕЧЕВОГО СИГНАЛА В СИСТЕМАХ СВЯЗИ

2.1 Вводные замечания

В основе развития и совершенствования устройств преобразования РС в системах телекоммуникаций лежит моделирование процессов обработки РС с целью выявления возможных направлений развития подобных систем.

В данном разделе диссертации развиты теоретические аспекты анализа и обработки РС в системах связи, представлены принципы построения систем его обработки на основе систем со структурно-параметрической адаптацией (подраздел 2.2), разработан метод выделения сегментов анализа РС различной длительности на однородных участках РС (подраздел 2.3), рассмотрены теоретические подходы к расчету параметров синтезирующей системы при обработке РС (подраздел 2.4), показаны технические решения по вычислению и коррекции линейных спектральных частот на однородных сегментах анализа при обработке РС (подраздел 2.5).

2.2 Обработка речевого сигнала на основе структурно-параметрической адаптации

Для создания перспективных алгоритмов обработки РС целесообразно использовать системы с переменной структурой, что позволит повысить качество их функционирования наряду с уменьшением скорости передачи, так как система, адаптивно меняя свою структуру, выбирает наилучшую из заданного конечного множества вариантов структур при ограничениях на скорость передачи [20, 21, 23, 24, 25, 68, 94, 192]. Анализ методов и алгоритмов обработки РС указывает на актуальность выполнения работ по созданию нового поколения систем подобного класса, основанных на исследовании статистических и параметрических характеристик распределения параметров РС с изменением в соответствии с ними структуры и параметров системы [128]. Такое техническое решение является новым

направлением развития систем обработки РС. Система обработки с переменной структурой состоит из следующих составных элементов:

- выделители параметров классификации;
- классификаторы;
- устройства, обеспечивающие реализацию макропроцедур обработки РС;
- устройства для дополнительной обработки РС.

Процессом изменения структуры и параметров устройства обработки в реальном масштабе времени управляет устройство, в котором происходит анализ РС и определение его параметров [24, 25, 119, 147].

В соответствии с вычисляемыми параметрами РС и их анализом выбирается одна из возможных структур обрабатывающего устройства, которая наиболее адекватно отобразит РС на сегменте (рисунок 2.1).

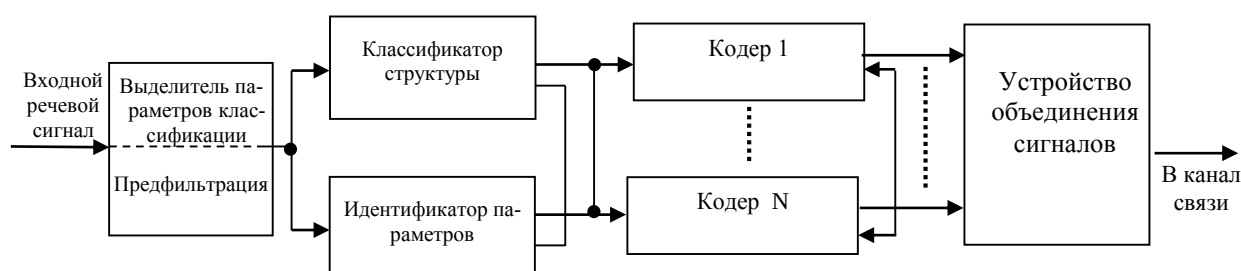


Рисунок 2.1 – Структура системы кодирования

Следующим шагом оптимизации системы является количественный и качественный выбор числа параметров, необходимых для обработки РС с заданным качеством. При обработке РС функционируют два адаптивных контура управления: первый производит оптимизацию структуры обрабатывающего устройства, второй – оптимизацию передаваемых параметров для выбранной структуры системы. Разработка систем подобного класса рассматривается как многокритериальная оптимизационная задача [20]. Подобный класс оптимизационных задач возникает тогда, когда оптимизация ведется по нескольким различным критериям, предъявляемым к системе компрессии [162]. При решении данной задачи целесообразно воспользоваться методами, описанными в [31, 65, 68, 94, 121, 173]. Если структура образуется набором элементов со связями между ними, то ее состояние удобно представлять двоичным вектором, определяющим ее построение:

$$\vec{W} = (w_1, \dots, w_k) \quad (2.1)$$

где w_k характеризует наличие ($w_k = 1$) или отсутствие ($w_k = 0$) k -го структурного элемента в системе. На вектор $\vec{W}$ могут быть наложены ограничения в виде равенств и неравенств, характеризующих зависимость между структурными элементами, тогда проектирование системы обработки может быть представлено следующей оптимизационной задачей:

$$\begin{aligned} S &\leq S_0; & A &\rightarrow \max; \\ A_n &\geq A_{0n}(P_c); & A &= \psi(K, \vec{W}); \\ W &\in [w_1, \dots, w_n]; & V_m &= \text{var}; V_m \in [V_{m1} \dots V_{mn}]; \\ & & \tau &\leq \tau_0; \end{aligned} \quad (2.2)$$

$$A(w) \rightarrow \max_{w \in N} \Rightarrow W_{0n},$$

где A – показатель качества, V_m – скорость передачи; S_0 – сложность ее функционирования; τ_0 – время задержки, P_c – вероятность ошибки в КС.

Данная задача может быть сведена к задаче математического программирования, решение которой при малых n реализуется полным перебором и методами, описанными в [173], а при больших – путем сведения к методам параметрического и/или непараметрического поиска.

Эволюционный метод формирования структуры системы предпочтителен в том случае, если проектируемая система W обработки может изменяться так, что будут соблюдаться накладываемые ограничения. При этом вариации δw структуры не должны выходить за пределы заданных ограничений, то есть:

$$w + \delta w \in N \quad (2.3)$$

Предложенные методы целесообразно применять в случае унимодальности задачи (2.2), однако это не всегда возможно для задач обработки РС, что требует дополнительных этапов анализа для поиска адекватного поставленным условиям решения. Основная трудность практической реализации таких систем связана с от-

сутствием априорной информации о вероятной мощности пространства представлений различных вариантов структур при оптимальном решении данной задачи [21, 22]. Устранение данной неопределенности частично реализуется при рациональном ограничении возможных вариантов при заданном требуемом качестве обработки, так как необходимо передавать информацию о состоянии и структуре системы в процессе ее функционирования. При исследовании систем обработки РС многие решения носят эмпирический характер, так как пока отсутствует четкий математический аппарат, который позволил бы связать воедино на основе компромисса между количеством возможных структур, качеством функционирования, алгоритмом работы и сложностью функционирования.

При построении математической модели необходимо учитывать особенности природы формирования РС в речевом аппарате человека. На рисунке 2.2 представлена предлагаемая обобщенная математическая модель системы обработки РС на основе структурно-параметрической адаптации, основным отличием которой от используемых в настоящий момент на практике является введение переменной длины сегмента анализа речевых данных.

В математической модели использованы операторы преобразований F_i , которые связывают между собой пространства представлений значений и параметров РС на различных этапах обработки [17].

При обработке РС, предполагая квазистационарный характер формирования сегментов, достаточно часто используется блочное представление речевых данных. Совокупность речевых отсчетов блока данных объединяется в сегмент анализа. В данной модели также используется сегментное представление данных. Под результатами обработки (множество $\{A_6\}$), в зависимости от цели использования модели, может выступать пространство обработанных речевых отсчетов на выходе низкоскоростного КРС, множество очищенных от шума и/или предварительно измененных в соответствии психоакустическим восприятием отсчетов речевых данных [136, 207, 225]. Технические решения, представленные в [136], являются новыми в данной предметной области.

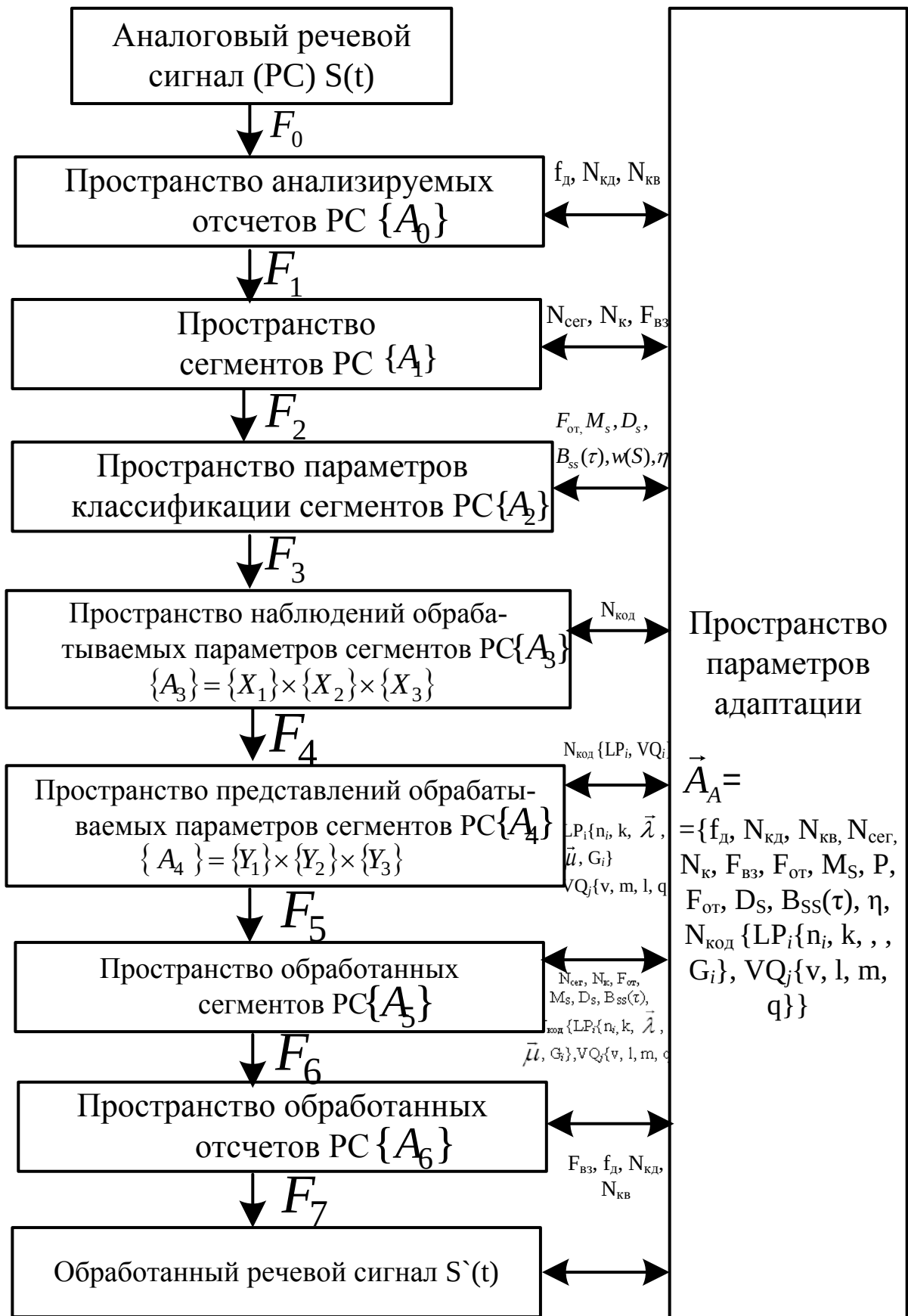


Рисунок 2.2 – Математическая модель системы обработки речевого сигнала

В рамках представленной модели также может быть реализована процедура идентификации говорящего на основе изменения функционального предназначения оператора F_6 . [157]. Результатами анализа РС при использовании модели являются пространства $\{A_2\}$, $\{A_3\}$, $\{A_4\}$, при реализации процедуры формирования пространств $\{A_5\}$, $\{A_6\}$.

Процесс преобразования описывается следующими выражениями:

1. $F_0 : \{s(t)\} \rightarrow \{A_0\}$. $S(n) \approx s(n\Delta t)$; $s(t) = \sum_{i=0}^N [R_i] \varphi_i(t)$, $\varphi_i(t) = \frac{\sin 2\pi f_{\text{верх}} (t - i/2 f_{\text{верх}})}{t - i/2 f_{\text{верх}}}$;
2. $F_1 : \{A_0\} \rightarrow \{A_1\}$. $S(n) = \bigcap_{k=0}^K (F(kl)S(kl))$; $l = 1 \dots N$, $n = 1 \dots n_{\text{max}}$, $K = (n_{\text{max}} \bmod N)$;
3. $F_2 : \{A_1\} \rightarrow \{A_2\}$. $A_2 = \{F_{\text{OT}} = 70 - 400 \text{ Гц}; \mu_s, D, B_{ss}, w_s, \eta, \theta\}$;
4. $F_3 : \{A_2\} \rightarrow \{A_3\}$. $\{A_3\} = \{N_{\text{код}}, LP_j \{k, \lambda, \mu, G, VQ_j(v, m, l, q)\}\}$;
5. $F_4 : \{A_3\} \rightarrow \{A_4\}$. $\{A_4\} = \vec{Q}\{N_{\text{код}}, LP_j \{k, \lambda, \mu, G, VQ_j(v, m, l, q)\}\}$;
6. $F_5 : \{A_4\} \rightarrow \{A_5\}$. $S'(n) = \bigcap_{k=0}^K (S'(kl))$; $l = 1 \dots N$, $n = 1 \dots n_{\text{max}}$, $K = (n_{\text{max}} \bmod N)$;
7. $F_6 : \{A_5\} \rightarrow \{A_6\}$. $A_6 = [S_k(1), S_k(2), \dots, S_k(n)]$;
8. $F_7 : \{\vec{A}_6\} \rightarrow \{S'(t)\}$.

При составлении математической модели применены обозначения:

$S(n)$ – отсчет РС;

$F(kl)$ – функция взвешивания;

Δt – интервал дискретизации;

$\varphi_i(t)$ – базисные функции ряда Котельникова;

K – количество сегментов; N – количество отсчетов в сегменте;

$[R_i]$ – коэффициенты базисных функций ряда Котельникова;

μ_s, D, B_{ss} – математическое ожидание, дисперсия и функция автокорреляции на сегменте анализа;

Q – операция квантования: скалярная или векторная;

ξ – вектор параметров ЛП;

η – количество переходов через ноль на сегменте анализа;

w_s – функция плотности вероятности (ФПВ) анализируемого сегмента;

Θ – количество классов сегментов;

$N_{\text{код}}$ – число кодирующих структур.

Параметры f_d – частота дискретизации аналогового сигнала, $N_{\text{кд}}$ – число бит кодирования, $N_{\text{кв}}$ – число уровней квантования, $F_{\text{вз}}$ – функция взвешивания, N_k – число отсчетов на сегменте анализа, $w(S)$ – типы функций плотности вероятности, $N_{\text{код}}$ – число подсистем кодирования берутся в дальнейшем фиксированными, так как одновременная перестройка всего спектра параметров адаптации потребует большого объема вычислений, выделения большого числа бит в цифровом потоке для их кодирования, что повлияет на алгоритмическую задержку и вычислительную сложность алгоритмов кодирования.

Основные этапы функционирования предложенной модели следующие:

1. $S(t)$ – пространство аналогового РС с ограниченными динамическим диапазоном и занимаемой полосой частот. В соответствии с рекомендациями МСЭ и особенностями разрабатываемого кодека:

$$f_{\text{нижн}} = 0,03 \text{ кГц}; \quad f_{\text{верх}} = 3,4 \text{ кГц}; \quad \text{Д (динамический диапазон)} = 43 \text{ дБ}.$$

Значение $f_{\text{нижн}} = 0,03 \text{ кГц}$ выбрано для того, чтобы провести анализ частоты ОТ.

2. F_0 – отображение пространства непрерывного сигнала $S(t)$ в пространство анализируемых отсчетов A_0 :

$$F_0: \{S(t)\} \rightarrow \{A_0\}. \quad (2.4)$$

Отображение F_0 описывает процедуру АЦП (дискретизацию, квантование, кодирование), частота дискретизации f_d , число уровней квантования $N_{\text{кв}}$, число бит кодирования $N_{\text{кд}}$ могут быть постоянным (например, по рекомендациям МСЭ $f_d = 8 \text{ кГц}$, $N_{\text{кв}} = 8$) или выбираться адаптивно. Математически F_0 может быть описано рядом Котельникова В. А. [166, 167]:

$$A(t) = \sum_{i=0}^N [R_i] \varphi_i(t), \quad (2.5)$$

где $[R_i]$ – квантованные и кодированные коэффициенты ряда Котельникова,

$$\varphi_i(t) = \frac{\sin 2\pi f_{\text{верх}} \left(t - \frac{i}{2f_{\text{верх}}}\right)}{t - \frac{i}{2f_{\text{верх}}}} - \text{базисные функции Котельникова,}$$

$f_{\text{верх}}$ – верхняя частота в спектре обрабатываемого РС.

$\{A_0\}$ – пространство анализируемых отсчетов:

$\{A_0\} = [S(1), S(2), \dots, S(n)]$, где $n \in \{1; R\}$, $R < \infty$, n – номер отсчета.

3. F_1 – отображение пространства анализируемых отсчетов $\{A_0\}$ в пространство анализируемых сегментов $\{A_1\}$.

$$F_1: \{A_0\} \rightarrow \{A_1\}.$$

В данном отображении N_k – число отсчетов в сегменте анализа. Обычно $N_k = 10\text{--}200$ отсчетов и может быть величиной переменной.

$N_{\text{сег}}$ – число сегментов, $N_{\text{сег}} < \infty$. $F_{\text{вз}}$ – функция взвешивания – функция, умножение на которую может обеспечить улучшение качества обработки РС.

$\{A_1\}$ – пространство анализируемых сегментов.

$\{\vec{A}_{11}\} = [S(1), S(2), \dots, S(N_k)]$, где $\vec{A}_{11} \in \{A_1\}$; $\vec{A}_{11}$ – анализируемый сегмент.

Пространство анализируемых сегментов РС $\{A_1\}$, образованное множеством возможных N -мерных векторов речевых отсчетов

$$\vec{S}_1 = [s(1), s(2), \dots, s(n), \dots, s(N)]^T; \vec{S}_1 \in \{S\}, \quad (2.6)$$

где N – количество отсчетов РС на интервале квазистационарности Δt (длительности анализируемого сегмента речи):

$$N = \Delta t \cdot f_{\partial}, \quad (2.7)$$

где f_{∂} – частота дискретизации РС (как правило, $f_{\partial} = 8$ кГц).

4. F_2 – отображение пространства анализируемых сегментов $\{A_1\}$ в пространство параметров классификации сегмента $\{A_2\}$. F_2 выполняет роль первого параметрического анализатора сегментов [6].

$$F_2: \{A_1\} \rightarrow \{A_2\}. \quad (2.8)$$

Способы реализации F_2 представлены в разделе 4.

$\{A_2\}$ – пространство параметров классификации сегментов РС.

Для реализации предлагаемых технических решений, представленных в разделе 6, $\{A_2\}$ принимает вид: $\{A_2\} = \{N_k, F_{от}, M_S, D_S, B_{SS}, \eta, P\}$, где: M_S – математическое ожидание амплитуды отсчетов на сегменте анализа при условии равновероятности появления отсчетов в сегменте; B_{SS} – функция автокорреляции РС на сегменте анализа; η – коэффициент огибающей кодируемого сигнала; P – средняя мощность сигнала на участке анализа.

5. F_3 – отображение пространства наблюдений обрабатываемых параметров сегментов $\{A_3\}$ на основе пространства параметров классификации сегментов $\{A_2\}$.

$$F_3: \{A_2\} \rightarrow \{A_3\}. \quad (2.9)$$

При низкоскоростном кодировании речи на основе модели ЛП пространство наблюдений кодируемых параметров $\{A_3\}$ представляет собой совокупность признаков пространств ЛП. Пространство наблюдений кодируемых параметров $\{A_3\}$ (признаковое пространство), образованное декартовым произведением подмножеств параметров модели и параметров сигнала ошибки предсказания анализирующего фильтра (сигнала возбуждения):

$$\{A_3\} = \{X_1\} \times \{X_2\} \times \{X_3\}. \quad (2.10)$$

В состав (2.10) входят:

$\{X_1\}$ – подмножество M -мерных векторов описывающих передаточную функцию голосового тракта;

$\{X_2\}$ – подмножество N -мерных векторов возбуждения;

$\{X_3\}$ – подмножество коэффициентов усиления векторов подмножества $\{X_2\}$ (коэффициентов G).

В рассмотренной модели считается, что множества $\{X_i\}$, $i = \overline{1,3}$ являются бесконечными и независимыми, что обусловлено природой РС и свойствами используемого параметрического анализа. Степень выявления и использования существующих зависимостей внутри множеств в значительной мере определяет эффективность параметрического липредерного кодирования речи тем или иным алгоритмом.

6. F_4 – отображение пространства наблюдений обрабатываемых параметров сегментов $\{A_3\}$ в пространство представлений обрабатываемых параметров сегментов $\{A_4\}$.

$$F_4: \{A_3\} \rightarrow \{A_4\}. \quad (2.11)$$

F_4 описывает процедуру кодирования параметров системы обработки РС.

Отображения подпространств $\{\vec{X}_i\}$ пространства наблюдений кодируемых параметров в подпространства $\{Y_i\}$ пространства представлений кодируемых параметров:

$$\{\vec{X}_i\} \rightarrow \{\vec{Y}_i\}. \quad (2.12)$$

Смысловым содержанием F_4 выступает процедура ВК параметров РС используемых при ЛП. Необходимо отметить, что при решении многих конкретных задач существует возможность объединения частных отображений в различных сочетаниях, вплоть до рассмотрения общего отображения

$$F_4 : \{\vec{X}\} \rightarrow \{\vec{Y}\}. \quad (2.13)$$

Пространство представлений кодируемых параметров $\{\vec{Y}\}$, образованное декартовым произведением подмножеств $\{Y_i\}$:

$$\{A_4\} = \{Y_1\} \times \{Y_2\} \times \{Y_3\}. \quad (2.14)$$

Соответственно, элементами множества $\{\vec{Y}\}$ являются векторы

$$\{\vec{Y}_i\} = (\vec{Y}_1, \vec{Y}_2, Y_3)^T. \quad (2.15)$$

Подмножества $\{Y_i\}$ являются конечными подмножествами определенной мощности, каждое из которых принадлежит множеству натуральных чисел $Y_i = \overline{1, y_i}$. Суммарная мощность пространства $\{\bar{Y}\}$ связана со скоростью кодирования текущего сегмента РС.

Общее число элементов данного множества составляет

$$\gamma = |\{Y_i\}| = \prod_{i=1}^3 y_i. \quad (2.16)$$

Скорость двоичного кодирования текущего сегмента РС определяется выражением

$$V = \log \prod_{i=1}^3 y_i = \log \gamma, \text{ бит/сегмент}, \quad (2.17)$$

которое позволяет определить область изменения $\gamma \leq 2^{V_k \cdot \Delta t}$, где V_k – определенная скорость кодирования.

7. F_5 – отображение пространства представлений кодируемых параметров $\{A_4\}$ в пространство оценок восстановленных сегментов РС $\{A_5\}$.

Отображение пространства представлений кодируемых параметров $\{\bar{Y}\}$ в пространство оценок анализируемых сегментов РС $\{\hat{\bar{S}}\}$:

$$F_5 : \{\bar{Y}\} \rightarrow \{\hat{\bar{S}}\}. \quad (2.18)$$

Отображение F_5 описывает процедуру восстановления сегментов РС. Как правило, нелинейный оператор F_5 реализует процедуры восстановления квантованных значений кодируемых параметров и получения (вычисления) оценок РС при его обработке. Восстановленный РС имеет погрешности, обусловленные наличием неточности отображений и конечности пространств представлений. F_5 – может также учитывать ошибки в КС.

Пространство оценок анализируемых сегментов РС $\{\hat{\bar{S}}\}$, изоморфное пространству анализируемых сегментов РС $\{\bar{S}\}$:

$$\hat{\bar{S}} = [\hat{s}(1), \hat{s}(2), \dots, \hat{s}(n), \dots, \hat{s}(N)]^T. \quad (2.19)$$

8. F_6 – отображение пространства обработанных сегментов РС $\{A_5\}$ в пространство обработанных отсчетов $\{A_6\}$.

$$F_6: \{A_5\} \rightarrow \{A_6\}. \quad (2.20)$$

Математически F_6 описывает формирование единого цифрового потока синтезированного РС. При соединении соседних сегментов на границе последние отсчеты восстанавливаются путем линейной интерполяции, перекрытия сегментов и применения оконных функций.

9. $\{A_6\}$ – пространство обработанных отсчетов РС

$$\{A_6\} = [S(1), S(2), \dots, S(n)]^T; \quad n \in \{1; R\}, \text{ где } R < \infty. \quad (2.21)$$

10. F_7 – отображение пространства обработанных отсчетов $\{A_6\}$ в пространство $S'(t)$ восстановленного аналогового РС :

$$F_7: \{A_6\} \rightarrow \{S'(t)\} \quad (2.22)$$

Отображение F_7 описывает процедуру ЦАП, частота дискретизации f_d , число уровней квантования $N_{\text{кв}}$ и число бит кодирования $N_{\text{кд}}$ могут быть переменными и выбираться адаптивно.

11. $S'(t)$ – пространство восстановленного аналогового РС с ограниченными динамическим диапазоном и полосой частот.

Формально существует еще одно отображение пространства анализируемых отсчетов $\vec{A}_0$ и пространства параметров адаптации $\vec{A}_a$ в пространство восстановленных отсчетов РС $\vec{A}_7$:

$$\{A_0, A_a\} \rightarrow \{A_6\}. \quad (2.23)$$

Это оператор связи между входным и выходным сигналами и параметрами адаптации. Нахождение аналитических зависимостей между этими тремя пространствами практически затруднено, так как при оценке многих характеристик и параметров РС используются субъективные неформализуемые в явном виде операторы.

Результатом исследований, направленных на нахождение параметров адаптации, в большей степени влияющих на качество кодирования РС, явилось нахождение достаточного и необходимого числа разделений классов сегментов РС, рациональных границ полос фильтрации, создание наборов КК параметров липредеров, формирование корректных КК ВК. Их описание представлено в 4 разделе.

12. Обобщенное пространство параметров адаптации A_a в соответствии с рисунком 2.2 имеет вид:

$$\vec{A}_a = \{f_d, N_{кд}, N_{кв}, N_{сег}, N_k, F_{вз}, F_{от}, M_S, B_{SS}, w(S), N_{код} \{LP_i \{N_k, n_i, T, \vec{\lambda}, \vec{\mu}, G_i\}, VQ_j \{N_k, v_j, l_j, q_j, M\}\}\}. \quad (2.24)$$

Пространство $\vec{A}_a$ описывает ресурсы, которые могут использоваться для процедуры адаптации.

Задача на создание системы обработки РС при низкоскоростном кодировании аналитически может быть представлена следующим образом:

$$V_m |_{A \geq A_0}; \xrightarrow{A'_a} \min \quad (2.25)$$

Использование полного перечня параметров адаптации нецелесообразно, так как влечет необоснованное увеличение вычислительной сложности системы обработки без заметного увеличения качества кодирования РС при заданной скорости передачи. Увеличение числа параметров адаптации может привести к увеличению доли служебной информации, передаваемой на приемную часть при низкоскоростном кодировании РС, что в свою очередь сократит информационный ресурс, выделяемый для передачи параметров подсистем кодирования. Оптимизационные задачи (2.25) с учетом вышеизложенного целесообразно упростить и привести к виду (2.26), где число параметров адаптации сокращается за счет выбора тех, которые оказывают доминирующее влияние на качество кодирования при адаптации структуры системы для заданных подсистем кодирования:

$$V_m |_{A \geq A_0}; \xrightarrow{A'_a} \min, \text{ при } \vec{A}'_a = \{N_{сег}, N_k, F_{вз}, F_{от}, M_S, B_{SS}, N_{код} \{LP_i, VQ_j\}\}. \quad (2.26)$$

Параметры f_d – частота дискретизации аналогового сигнала, $N_{кд}$ – число бит кодирования, $N_{кв}$ – число уровней квантования, $F_{вз}$ – функция взвешивания, N_k –

число отсчетов на сегменте анализа, $w(S)$ – типы ФПВ, $N_{\text{код}}$ – число подсистем кодирования берутся в дальнейшем фиксированными, так как одновременная перестройка всего спектра параметров адаптации потребует большого объема вычислений, выделения большого числа бит в цифровом потоке для их представления. Это влияет на алгоритмическую задержку и вычислительную сложность алгоритмов. Следует отметить, что разрабатываемые системы обработки предназначены для сопряжения с существующими системами АЦП и ЦАП, что накладывает определенные ограничения на их разработку. Параметры адаптации, приведенные в (2.26), вычисляются на каждом сегменте анализа, а их учет при обработке может обеспечить существенное повышение качества кодирования РС [22, 41, 44, 120, 157]. Существующие системы обработки РС используют постоянную структуру системы обработки, что негативно сказывается на возможности учитывать особенности РС на различных сегментах анализа. Задача (2.26) решается путем ее декомпозиции в два этапа: на первом из них осуществляется поиск рациональной (оптимальной) структуры системы, а на втором вычисляются оптимальные параметры для выбранного варианта системы обработки.

Разработанная математическая модель системы с переменной структурой для обработки РС с учетом переменной длины сегмента анализа дала возможность по-новому сформулировать задачу разработки алгоритмов низкоскоростного кодирования РС. Таким образом, к задачам разработки КРС, приведенным в первой главе, добавился учет возможности введения структурной адаптации системы за счет априорной и апостериорной информации о РС на этапе анализа при переменной длительности сегмента анализа.

Анализ задачи (2.26) позволяет классифицировать ее как задачу формирования системы обработки со структурно-параметрической адаптацией. Так, поиск рационального (оптимального) числа систем субкодексов $N_{\text{код}}$, номеров субкодексов LP_i , VQ_j , участвующих в обработке конкретного субсегмента РС, N_k – количество отсчетов на сегменте анализа направлены на поиск рациональной структуры системы обработки для конкретного класса сегментов переменной длительности.

Задача выбора структуры системы решалась в соответствии с результатами анализа свойств РС, принципов построения существующих КРС, методов структурной адаптации, применяемых в других областях науки и техники.

Поиск оптимальных ВК наборов КЛП осуществляется процедурой «анализ через синтез», описанной в [188]. На текущем сегменте анализа элементы пространства анализируемых сегментов РС представлены вектором $\vec{S}_0$ (2.26). Его отображением является вектор $\vec{X}_0$ пространства наблюдений кодируемых параметров: $F_3 : \vec{S}_0 \rightarrow \vec{X}_0$. В пространстве наблюдений кодируемых параметров зафиксированы вектора $\vec{X}_{10}$, $\vec{X}_{20}$ и скаляр X_{30} . Данные значения признаков в результате действия F_4 отобразятся в конкретные значения скаляров Y_i , которые в совокупности определяют конкретный вектор пространства представлений кодируемых параметров $\vec{Y} = F_4([\vec{X}_i], k, l, m)$. Результатом отображения вектора $\vec{Y}$ в пространство оценок анализируемых сегментов РС на основании (2.18) является вектор $\hat{\vec{S}} = F_5(\vec{Y})$.

$$D(\vec{S}_0, \hat{\vec{S}}_0) = D[\vec{S}_0, F_5(\vec{Y})] \quad (2.27)$$

Очевидно, что результат выражения (2.26) является функцией случайного вектора $(\vec{X}_{10}, \vec{X}_{20}, X_{30})$, отображаемого в вектор $\vec{Y}$. Координаты данного случайного вектора считаются независимыми, следовательно, их многомерная совместная плотность распределения имеет вид

$$W(\vec{X}_{10}, \vec{X}_{20}, X_{30}) = W_1(\vec{X}_{10}) \cdot W_2(\vec{X}_{20}) \cdot W_3(X_{30}). \quad (2.28)$$

С учетом случайного характера вектора $(\vec{X}_{10}, \vec{X}_{20}, X_{30})$ усреднение величины искажения РС по всем анализируемым речевым векторам, отображаемым в вектор $\vec{Y}$ определяется следующим образом:

$$M[D(\vec{S}_0, \hat{\vec{S}}_0)] = \int_{\{\vec{X}_{1_0}\}} \int_{\{\vec{X}_{2_0}\}} \int_{\{\vec{X}_{3_0}\}} D(\vec{S}_0, \hat{\vec{S}}_0) W(\vec{X}_{1_0}, \vec{X}_{2_0}, X_{3_0}) d\vec{X}_{1_0} d\vec{X}_{2_0} dX_{3_0} \quad (2.29)$$

При изменении вектора $\vec{Y}$ результат расчета (2.28) представляет собой дискретную случайную величину.

Квантующие отображения F_4 заключаются в разделении бесконечных множеств $\{\vec{X}_i\}$ на конечные числа y_i непересекающихся классов $\{\vec{R}_{ij}\}$, $j = \overline{1, y_i}$, каждому из которых соответствует свой элемент из $\{Y_i\}$:

$$\begin{cases} \{\vec{R}_{i1}\} \cup \{\vec{R}_{i2}\} \cup \dots \cup \{\vec{R}_{iy_i}\} = \{\vec{X}_i\} \\ \{\vec{R}_{ik}\} \cap \{\vec{R}_{ij}\} = 0 \text{ при } k \neq j; k, j \in \overline{1, y_i} \\ F_{4_i} : \{\vec{R}_{ij}\} \rightarrow Y_{ij} \end{cases} \quad (2.30)$$

На основании (2.30) можно записать полученное искажение РС в виде:

$$D(\vec{S}_0, \hat{\vec{S}}_0) = \sum_{y_1=1}^{Y_1} \sum_{y_2=1}^{Y_2} \sum_{y_3=1}^{Y_3} D[\vec{X}_0, F_4([X_i], k, l, m)] \cdot P_w(\vec{X}_{1_0}, \vec{X}_{2_0}, X_{3_0}). \quad (2.31)$$

Полученная величина $D(\vec{S}_0, \hat{\vec{S}}_0)$ характеризует среднее значение избранной меры искажения РС при обработке последнего алгоритмом гибридного кодирования на основе метода ЛП.

Данная модель является основой для разработки математической модели кодирования РС, учитывающей зависимости элементов декомпозиции речи, поскольку в рассмотренной обобщенной математической модели векторы описывающие параметры голосового тракта и сигнала возбуждения трактуются как независимые, что является существенным недостатком и требует увеличения информационного ресурса на представление сигнала возбуждения при обработке РС.

Вновь разработанная модель должна удовлетворять следующим принципам: адекватность; соответствие модели решаемой задаче; упрощение при сохранении существенных свойств системы; соответствие между требуемой точностью результатов моделирования и сложностью модели; баланс погрешностей различных видов; многовариантность реализаций элементов модели; блочное строение.

1. Адекватность.

Разработанная модель по основным параметрам и свойствам соответствует стандартам МСЭ G.711, G.726, G.728, G.729, а также [83, 84].

По сложности модель системы сопоставима с кодеком стандарта G.729, новым в ней является оригинальный способ структурной адаптации [128], не требующий существенного усложнения системы обработки.

2. Соответствие модели решаемой задаче.

Модель разрабатывалась исходя из анализа следующих факторов: тенденций развития связи России, тенденций развития КРС, требований, предъявляемых к КРС МСЭ и стандартами России.

3. Упрощение при сохранении существенных свойств системы.

Основные требования к КРС в модели учтены, а в качестве упрощений не учитываются конечная разность КЛП липредеров и ошибки квантования ИКМ.

4. Соответствие между требуемой точностью результатов моделирования и сложностью модели.

В модели учтены все переменные, используемые в известных моделях, особенности параметров модели и функциональные связи между переменными не изменялись.

5. Баланс погрешностей различных видов.

Погрешности моделирования обусловлены точностью представления входных и выходных данных, точностью элементов системы, случайные погрешности являются величинами одного порядка и поэтому влияют на точность обработки равноценно.

6. Многовариантность реализаций элементов модели.

Основные элементы модели являются многовариантными. Так, в качестве кодеков используются наборы липредеров с ВК и различными наборами КК, определяемыми в процессе обучения.

7. Блочное строение.

В системе можно выделить следующие подсистемы: предварительной обработки; многоуровневый классификатор, наборы кодеков сегментов (липредеры с ВК); подсистемы объединения и разделения сигналов; наборы декодеров; восстановления сигнала.

2.3 Метод деления речевого сигнала на сегменты различной длительности

В основной массе рекомендаций по низкоскоростному кодированию РС используется подход, основанный на определении значений параметров модели речеобразования путем кратковременного анализа сегментов речи фиксированной длительности от 10 до 30 мс [127, 146]. Фиксированный сегмент анализа речевых данных используется в большинстве систем обработки и кодирования РС, что является существенным недостатком данных устройств в условиях перехода к системам с пакетной коммутацией и переменной скоростью кодирования. Формирование переменной длины сегмента анализа РС основано на использовании долговременных корреляционных зависимостей присутствующих в РС и взаимосвязью между функцией автокорреляции РС на анализируемом сегменте $B(\tau)$ и его спектральной плотностью мощности $G(\omega)$. Данное техническое решение представлено в [137] и является основой для выделения сегментов анализа РС различной длительности.

Представленная пара преобразований позволяет определить спектральную плотность мощности по автокорреляционной функции случайного РС на основании теоремы А. Я. Хинчина и Н. Винера [166, 168]. При этом необходимо учитывать требования стационарности для случайного РС. В соответствии с существую-

шей связью между $B(\tau)$ и $G(\omega)$ можно определить другие важные показатели, характеризующие случайный РС. К ним можно отнести интервал корреляции $\tau_{кор}$, характеризующий промежуток времени для случайного РС, мгновенные значения которого взаимосвязаны, следовательно, они имеют одну структуру формирования и эффективную полосу частот $F_{эф}$, определяющую частотный диапазон в котором сосредоточено 90 – 95 % мощности. Существует взаимосвязь между $\tau_{кор}$ и $F_{эф}$. Эта взаимосвязь устанавливается через произведение этих параметров [98].

Значение $\tau_{кор}$ определяется точкой первого перехода автокорреляционной функции через ноль [98, 168]. Математические модели и установленные соотношения между $\tau_{кор}$ и $F_{эф}$ позволяют использовать их для анализа реального РС. Для РС слитной русской речи, помимо возможности классификации активного участка на тоновой или шумовой [28, 89, 129, 142] важной характеристикой является эффективная полоса частот $F_{эф}$ занимаемая речью. На рисунках 2.3, 2.4 представлены временное представление, автокорреляционная функция и спектральное представление звуков "А" и "Ш" соответственно. Для оценки спектральной плотности мощности (СПМ) звуков "А" и "Ш" использовались тестовые фразы, которые взяты из [84], их перечень и особенности формирования представлены в Приложении 1.

Анализ графиков усредненных значений кратковременной спектральной плотности мощности РС показывает, что в полосе частот от 3 до 7 кГц появляются всплески спектральной энергии только при произнесении некоторой части согласных звуков (например, "ш", "с", "ф", "х"), которые в речи встречаются относительно редко [186]. Таким образом, в приложениях обработки РС увеличение частоты дискретизации от 8 до 16 кГц не приводит к существенному выигрышу в качественных показателях функционирования, однако значительно увеличивает количество обрабатываемых отсчетов и скорость передачи в КС при применении классических технических решений. Анализ и учет усредненных значений кратковременной спектральной плотности мощности РС позволил разработать технические решения по адаптации системы обработки, устраняющие данные недостатки [8, 134].

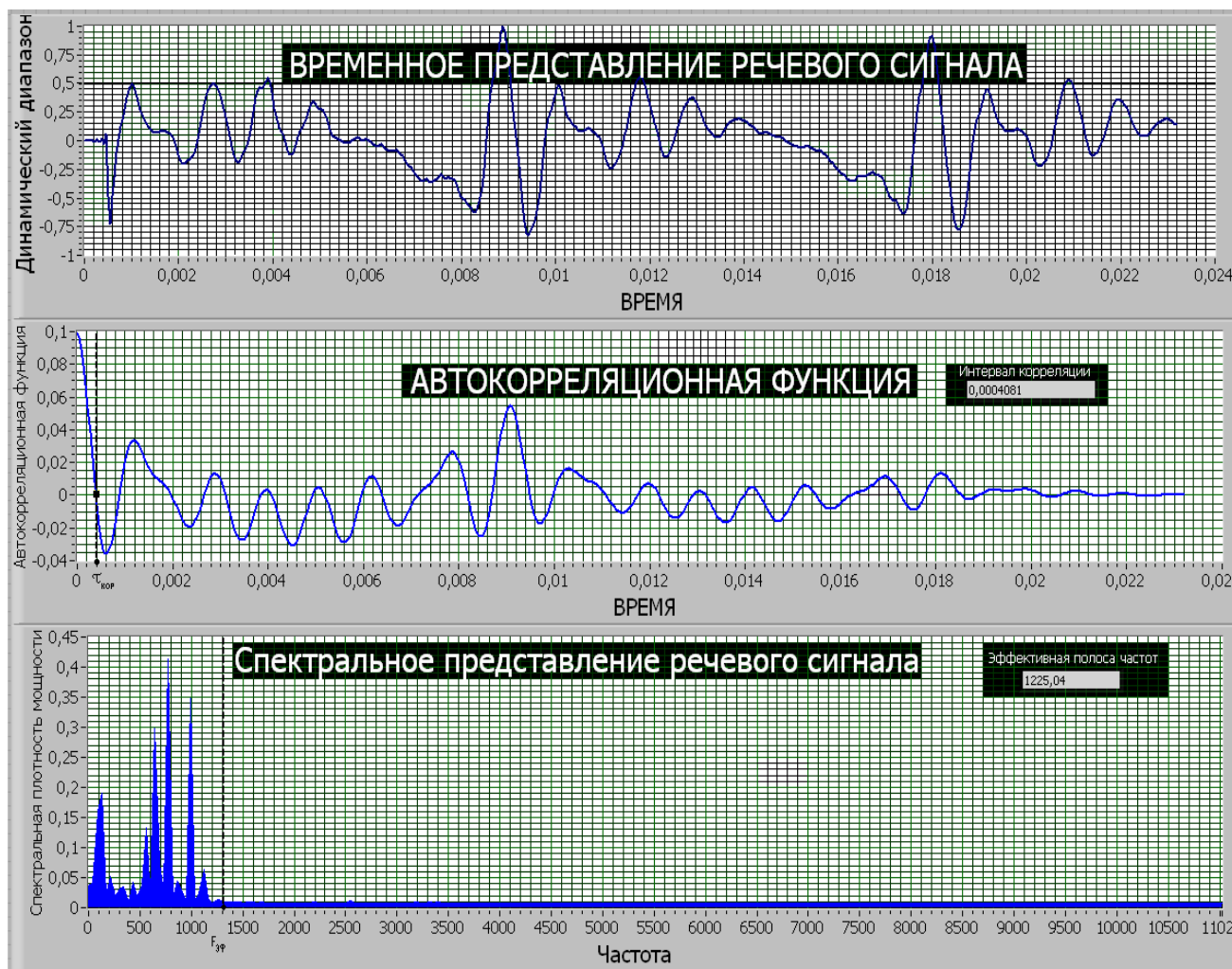


Рисунок 2.3 – Временное представление, автокорреляционная функция и спектральное представление звука "А"

Чередование звуков слитной речи связано с изменением интервала корреляции, который указывает на нарушение режима однородности сегмента РС. Выбор длительности сегментов анализа данных является весьма важной задачей при исследовании динамики изменения характеристик РС методом кратковременного анализа, в процессе которого оцениваются энергия, текущий спектр, число пересечений нуля и некоторые другие параметры [69, 188].

Следует отметить, что для различных звуков формируемой речи длительности устоявшегося процесса излучения значительно превышают 10 ... 30 мс, а длительность отдельных звуков речи составляет от 20 до 350 мс (таблицы 1.3, 1.4). Гласные звуки имеют большую длительность, чем согласные. Темп речи может изменяться в широких пределах, причем длительность гласных звуков изменяется в

большей степени. Максимально возможный интервал одновременно анализируемых данных составляет 60–80 мс, что связано с требованиями по задержке РС при его передаче, определяемыми рекомендацией G.114 Международного союза электросвязи. Таким образом, если на протяжении 60–80 мс не происходит смены сегмента однородности при формировании РС, то новый анализ начинается при использовании данных о выделенных параметрах предыдущего сегмента.

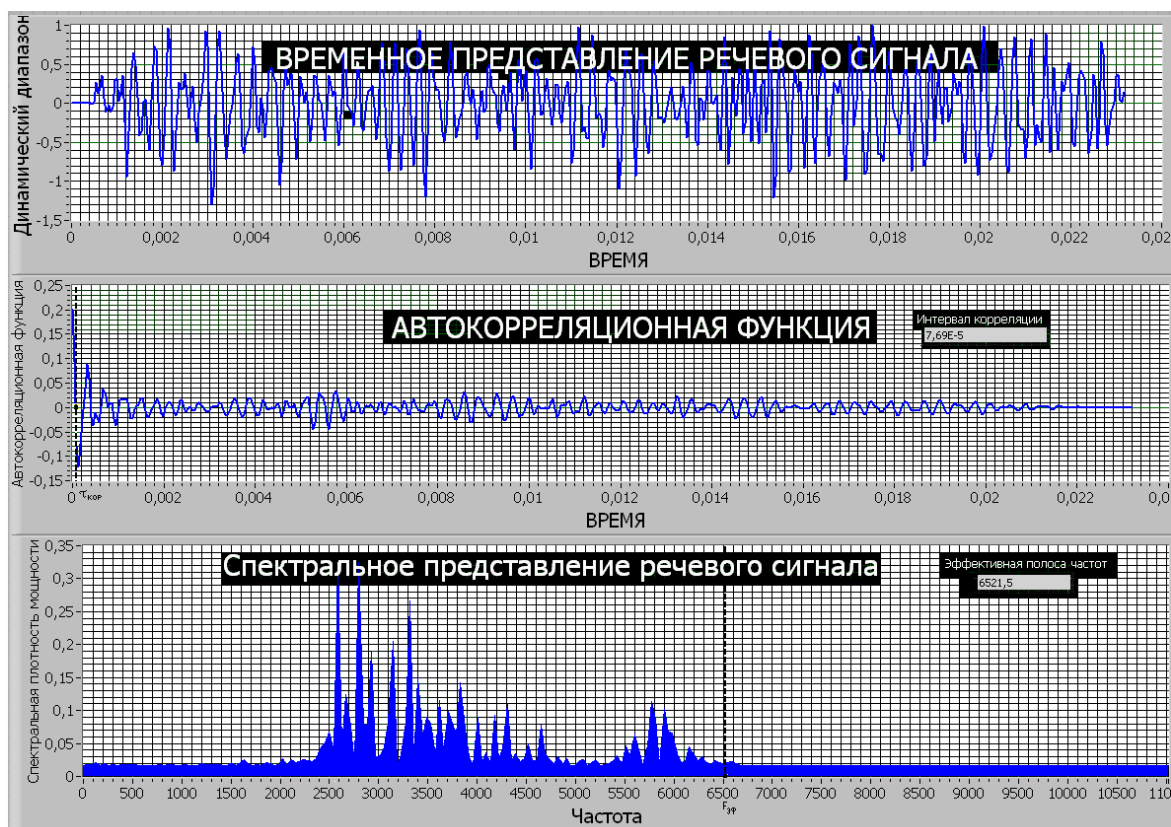


Рисунок 2.4 – Временное представление, автокорреляционная функция и спектральное представление звука "Ш"

Использование такого подхода к формированию сегментов обработки РС позволяет выделять однородные сегменты анализа [18]. Увеличение длительности кратковременного анализа приводит к тому, что параметры формирующей (передаточной) функции системы обработки и сигнала возбуждения будут сохраняться на всем протяжении анализа, что приводит к сокращению средней скорости передачи РС [9]. В качестве примера рассмотрим вокализованный сегмент данных на интервале нескольких периодов ОТ (рисунок 2.5). Одним из принципиальных моментов

при формировании конечной границы интервала анализа речевого сегмента является вычисление периода ОТ и формирование границ сегмента анализа в соответствии с его значением. Вычислением данного параметра производится в соответствии с [129]. При этом границы сегмента формируются на основании выражений (2.32)-(2.34).

$$\Delta T \in (t_n \dots t_k) = \text{var}; \Delta T > 20 \text{ мс}; \frac{\Delta T}{T_{\text{от}}} \in \{N\}; \quad (2.32)$$

$$t_n = \Delta t n, \text{ при } \text{sign}(S(t_n)) + \text{sign}(S(\Delta t(n-1))) = 0; \quad (2.33)$$

$$t_k = \Delta t k, \text{ при } \text{sign}(S(t_k)) = \text{sign}(S(t_n)) \& \text{sign}S(\Delta t(k-1)) = \text{sign}(S(\Delta t(n-1))) \& (t_k - t_n) > 20 \text{ мс}, \quad (2.34)$$

где ΔT – длительность сегмента анализа; t_n , t_k – время начала и окончания сегмента анализа; $T_{\text{от}}$ – период ОТ; n – номер отсчета в начале сегмента; k – номер отсчета в конце сегмента; Δt – интервал дискретизации, N – область натуральных чисел.

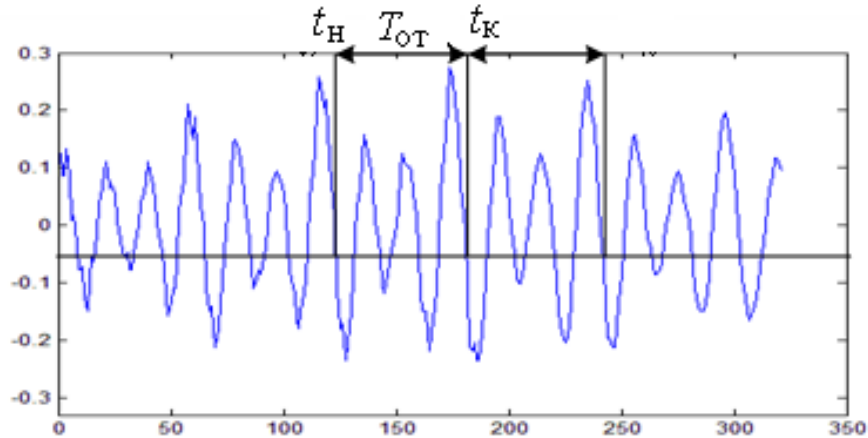


Рисунок 2.5 – Динамическое выделение участка квазистационарности

Использование выражений (2.32)–(2.34) позволяет получить сегменты, отсчеты которых в начале и окончании будут иметь одинаковые знаки конечной разности первого порядка, такой подход является новым направлением при выделении сегментов анализа и представлен в [133]. При этом с высокой вероятностью можно

утверждать, что начальный и конечный отсчеты во вновь сформированном сегменте будут иметь значения, близкие к моменту изменения полярности знака отсчетов РС, что значительно уменьшит возможные искажения на стыках сегментов. Такое техническое решение описано в [130] и является новым направлением развития систем анализа.

На шумоподобных сегментах формирование границы окончания сегмента анализа реализуется в соответствии с вышеизложенным подходом на интервалах соответствующих длительности около 20 мс. Таким образом, возможно использование более длительных интервалов анализа, особенно на сегментах имеющих квазипериодическую вокализованную природу образования. Возможность увеличения длины анализируемого сегмента связана с наличием корреляционных связей между отсчётами РС, которые выявляются с помощью анализа автокорреляционной функции [85, 95]. Однако применение данных лишь о корреляционных связях между отсчётами не полностью устраняет статистическую избыточность РС при кодировании, при этом может возникать ошибка в определении участков однородности при переходе между звуками одного фонетического ряда. В задачах прикладного характера распределение значений РС априорно является случайным, более того сам вид распределения остаётся неизвестным.

Увеличение границ анализируемого сегмента может быть основано на использовании более полной информации о статистических взаимосвязях между мгновенными значениями РС, получаемыми с помощью непараметрических методов, то есть при отсутствии данных о характере распределения. При адаптивном изменении границ сегмента анализа РС, количество наблюдаемых отсчётов, составляющих данный участок, априорно является неизвестной величиной. Оно может принимать различные значения в зависимости от особенностей произносимой речи. Соответственно количество наблюдений (отсчётов) составляющих анализируемый участок зависит в данном случае от исхода самих наблюдений. Следовательно, для определения длительности сегмента анализа необходимо применять методы последовательной проверки статистических гипотез, суть которых сво-

дится к определению "эффективных" выборок при которых будет иметь место расширение анализируемого сегмента, и, как следствие, увеличение набора отсчетов. В данном случае предлагается метод, основанный на последовательном критерии отношения вероятностей (критерий Вальда), которая заключается в вычислении отношения вероятностей получения выборок (функций правдоподобия) на каждом этапе эксперимента [72]. В силу отсутствия информации о характере (виде) распределения функция плотности вероятности, необходимая для вычисления правдоподобия, оценивается методом непараметрического ядерного сглаживания Парзена [71]:

$$f_{\text{я}}(X) = \sum_{i=1}^N K\left[\frac{s}{h} - \frac{s_i}{h}\right], \quad (2.35)$$

где $f_{\text{я}}(X)$ – ядерная оценка плотности, $K(u)$ – ядерная функция (окно), h – ширина окна.

В качестве ядерной функции используется гауссов профиль [104]:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{u^2}{2}\right)}, \quad (2.36)$$

Количество окон сглаживания для каждого начального сегмента анализа напрямую зависит от их ширины и является статистически обоснованным и определяемым с помощью информационного критерия Акаике (AIC) [191]. Критерий использует соотношение между относительной информацией (расстоянием) Кульбака–Лейблера (между истинной неизвестной функцией распределения и ее приближенным видом) и логарифмической функцией правдоподобия:

$$AIC = -2\ln(L(S, m)) + 2k, \quad (2.37)$$

где $L(S, m)$ функция правдоподобия для анализируемого набора отсчётов, зависящая от числа окон (ширины окна), k – количество параметров (в данном случае число окон, m).

Решение принимается в пользу минимального значения AIC . Получаемая в результате ядерного сглаживания кривая плотности распределения начального сегмента анализа $f_0(S, m)$, является основой для вычисления значения функции правдоподобия, в случае справедливости гипотезы H_0 (основной гипотезы), для последующих наборов отсчётов, получаемых на каждом этапе увеличения его длительности:

$$L_0(S, m) = \prod_{i=1}^N f_0(s_i, m), \quad (2.38)$$

где N – количество отсчётов составляющих анализируемый сегмент, $f_0(s_i, m)$ – значение плотности вероятности в точке s_i .

В случае попадания значения отсчета s_i , между значениями $f_0(S, m)$, вычисление значения плотности в данной точке $f(s_i)$ производится с помощью линейной интерполяции. Выражение (2.37) с учётом (2.38) и изменённым знаком, является основой для определения оптимального числа окон ядерного сглаживания:

$$m_{opt} = \arg \max_m \left[\ln \left(\prod_{i=1}^N f_0(s_i, m) \right) - m \right], \quad (2.39)$$

Определённое таким образом, оптимальное число окон сглаживания используется для оценивания функции плотности распределения при справедливости гипотезы H_1 (альтернативной) на каждом этапе эксперимента.

Основная гипотеза, в данном случае имеет вид: альтернативная $H_1: F(S, m) = F_1(S, m)$. При каждом увеличении сегмента на длительность периода ОТ на основе вновь полученного набора отсчётов РС производятся следующие операции:

1. Вычисляется значение правдоподобия при справедливости гипотезы H_0 по соотношению (2.38).
2. Производится ядерная оценка плотности распределения $f_1(s_i, m)$, с помощью (2.35).
3. Аналогично (2.38) производится вычисление значения правдоподобия $L_1(S, m)$ по $f_1(s_i, m)$.

Далее вычисляется статистика критерия Вальда [72]:

$$\ln \frac{L_1(S, m)}{L_0(S, m)} = \ln(L_1(S, m)) - \ln(L_0(S, m)) = Z[S], \quad (2.40)$$

При этом анализируемые выборки (сегменты) принадлежат к «эффективной» области $C_{эф}$ пространства выборок при выполнении:

$$\ln B|_{n < n'} < Z[S] < \ln A|_{n < n'}, \quad (2.41)$$

где n – общее количество отсчётов в анализируемом сегменте, n' – максимально возможное число отсчётов в сегменте, с учётом ограничений G.114 МСЭ.

В случае:

$$Z[S] < \ln B|_{n < n'}, \quad (2.42)$$

набор отсчётов принадлежит к "абсолютно эффективной" области $C_{эф}^{abc}$ (сегменту анализа). Оставшийся вариант при

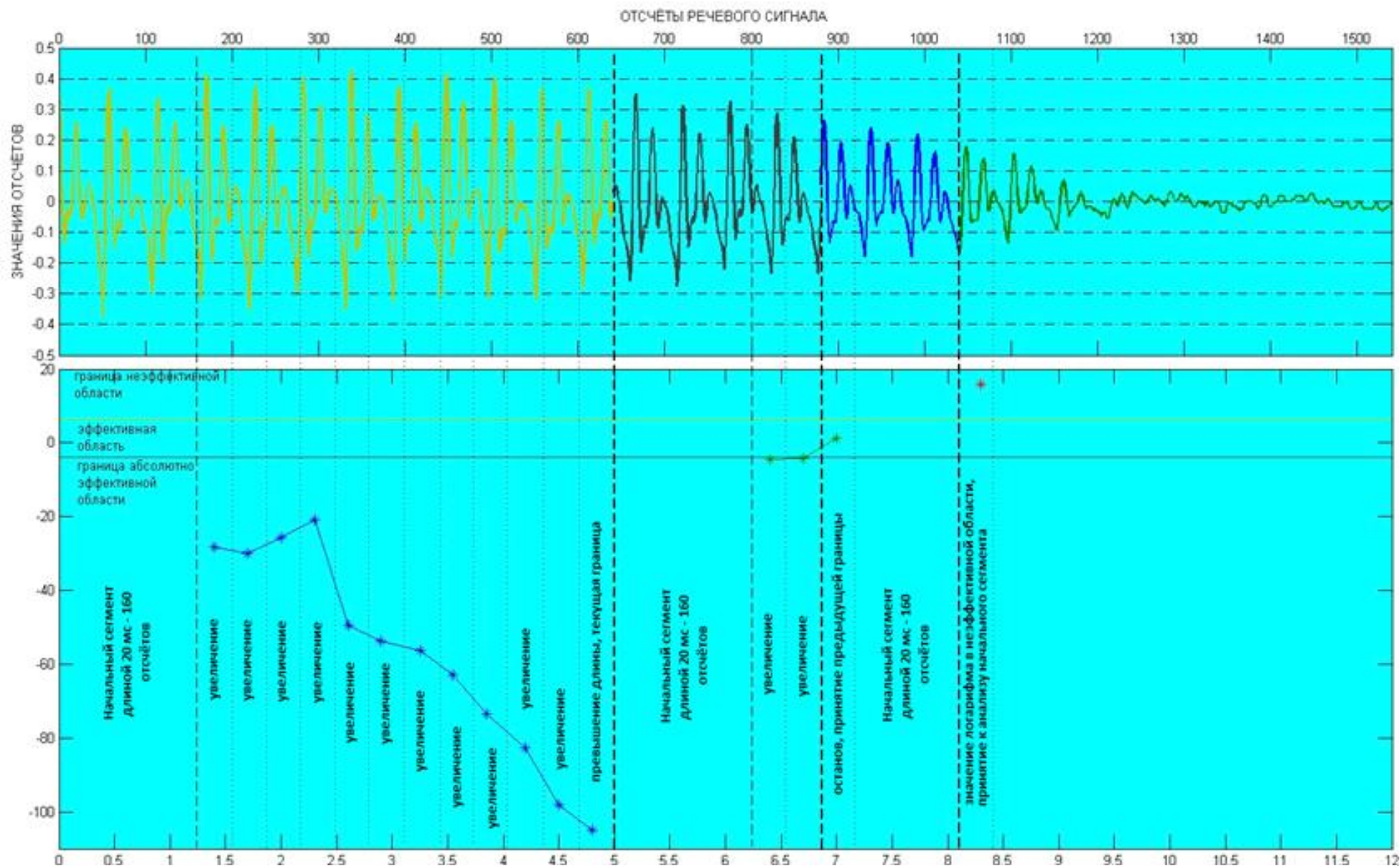
$$Z[S] > \ln A, \quad (2.43)$$

определяет набор отсчётов принадлежащим к «неэффективной» области $C_{н.эф}$ (сегменту анализа). Ограничительные константы A и B , определяются на основании ошибок первого α и второго β рода:

$$\begin{cases} \ln A = \ln \frac{1-\beta}{\alpha} \\ \ln B = \ln \frac{\beta}{1-\alpha} \end{cases}, \quad (2.44)$$

где $\alpha = 0,002$ и $\beta = 0,001$.

Анализ функции отношения вероятностей позволяет выделить важный показатель, характеризующий РС. Это интервал, являющийся промежутком времени, в течение которого отсчеты статистически взаимосвязаны, следовательно, являются участками однородности при формировании РС. Данный факт соотносится с результатами анализа образования вокализованных и шумоподобных сигналов [69, 186]. На рисунке 2.6 и рисунке 2.7 представлены временное представление и значения отношения вероятностей при сдвиге сегмента анализа на 5 мс, относительно приходящего РС при переходе с шумового на вокализованный (рисунок 2.6) участок и наоборот (рисунок 2.7).



Поведение статистики Вальда (логарифма отношения вероятностей) при последовательном увеличении сегмента в 160 мс на 5 мс

Рисунок 2.7 – Выделение сегментов обработки на основе последовательного статистического анализа (переход от вокализованного к шумовому сегменту)

Данные рисунки получены при произношении звуков слитной речи с использованием программного обеспечения, сформированного с использованием программы технических расчетов MATLAB [26, 30, 32, 88, 93, 241]. Аналого-цифровое преобразование РС выполнено с частотой дискретизации 8 кГц при использовании 8 битного линейного квантователя. Анализ (рисунка 2.6) и (рисунка 2.7) позволяет сделать вывод о том, что на однородных сегментах образования РС значения логарифма отношения вероятностей лежат в области либо принятия гипотезы H_0 , либо продолжения эксперимента (увеличения длительности сегмента анализа). При этом сегмент анализа берется длиной 20 мс, расширение сегмента осуществляется на 5 мс от значения начальной границы сегмента при сохранении знака конечной разности первого порядка, при этом граница сегмента анализа формируется в момент изменения однородности образования РС.

Критерием принятия решения о смене однородности при формировании сегментов РС является последовательный критерий отношения вероятностей Вальда, в основе которого лежит разделение m -мерного пространства отсчётов РС на три непересекающиеся области. При анализе речевых данных решение о продлении наблюдений, то есть о расширении границы сегмента анализа принимается в случае принадлежности анализируемой выборки двум из трёх областей пространства, "эффективной" либо "абсолютно эффективной". Решение о прекращении увеличения длительности сегмента принимается в случае принадлежности выборки к "неэффективной" области, либо переходе между "абсолютно эффективной" и "эффективной" областями. При этом в случае перехода из "эффективной" в "абсолютно эффективную" область граница сегмента сохраняется на данном этапе. В случае же перехода из "абсолютно эффективной" в "эффективную" область граница сегмента определяется по данным предыдущего этапа.

Максимально возможный сегмент одновременно анализируемых данных составляет 80 мс, что связано с требованиями по задержке РС при передаче, определяемыми рекомендацией G.114 Международного союза электросвязи. Если на протяжении 80 мс не произошло перехода между областями, то увеличение длитель-

ности сегмента прекращается. Экспериментально установлено, что принадлежность значения логарифма отношения "эффективной" области, соответствует переходному процессу, и в случае нахождения анализируемой выборки в «эффективной» области по достижении 80 мс процесс заканчивается принятием основной гипотезы. Показателем принадлежности сегмента анализа, получаемого на каждом этапе эксперимента, к определённой области является логарифм отношения функций правдоподобия $Z[X]$, вычисляемый по соотношению (2.40).

Использование такого подхода к формированию сегментов обработки РС позволяет наиболее полно использовать статистическую взаимосвязь между случайными значениями РС для выделения однородных сегментов анализа, что дает возможность не осуществлять на длительном сегменте анализа РС перестройки параметров формирующей (передаточной) функции голосового тракта и приводит к значительному сокращению средней скорости передачи РС при сохранении качественных показателей синтезированной речи [37].

Предлагаемое деление РС на однородные сегменты достаточно хорошо соотносится с квазистационарным характером РС на временных интервалах, соответствующих режиму установившихся звуков. Вычисляемые в ходе исследований значения линейных спектральных частот (ЛСЧ) позволяют утверждать о высокой корреляции между параметрами, описывающими значения максимумов амплитудно-частотной характеристики речевого тракта на близлежащих однородных сегментах РС, что может быть использовано при обработке РС в системах низкоскоростного кодирования речевых данных с переменной скоростью передачи [132]. Изменение длительности сегментов анализа РС в вокодерах с ЛП дает возможность уменьшить среднюю скорость передачи в УОРС, функционирующих с переменной скоростью передачи.

2.4 Определение параметров синтезирующего фильтра при кодировании речевого сигнала

Как было показано в разделе 1, среди многообразия методов кодирования РС наиболее эффективным является метод ЛП, который принадлежит к классу методов, использующих модель РС в виде отклика линейной системы с переменными параметрами (голосового тракта) на соответствующий сигнал возбуждения (порождающий сигнал). Анализатор речепреобразующего устройства выделяет из короткого сегмента РС параметры состояния линейной системы и сигнала возбуждения, позволяющие синтезатору восстановить исходный сигнал с требуемой степенью верности [145].

При этом амплитудно-частотная характеристика синтезирующего фильтра вычисляется согласно выражению.

$$A(\omega T) = \frac{1}{\sqrt{\left(1 + \sum_{m=1}^{M-1} a_m \cos m\omega T\right)^2 + \left(1 + \sum_{m=1}^{M-1} a_m \sin m\omega T\right)^2}}; \quad (2.45)$$

Исследования в предметной области речевого кодирования показали невозможность прямого интерполирования и квантования КЛП ($\vec{a}$) при переходе от одного субсегмента к другому в рамках единого однородного сегмента данных, так как это приводит к высокой вероятности получения неустойчивого фильтра синтеза модели предсказания [91, 109, 186].

При этом амплитудно-частотная характеристика фильтра синтеза весьма чувствительна к точности представления коэффициентов $\{a_k\}$ (их квантованию), поэтому для адекватного отображения состояния фильтра синтеза требуется: увеличивать точность квантования (разрядность) коэффициентов; использовать мощные помехозащитные коды (что также увеличивает необходимую разрядность представления коэффициентов); повышать качество КС; обеспечивать устойчивость системы обработки РС.

Невозможность передачи по КС КЛП породила необходимость поиска других параметров авторегрессии, которые являлись бы математическими эквивалентами КЛП и при этом обеспечивали:

- низкую чувствительность к квантованию и несущественное влияние на точность восстановления КЛП;
- возможность контроля устойчивости фильтра синтеза с минимальными вычислительными затратами;
- относительную простоту вычислений.

Для описания передаточной функции синтезирующего фильтра на практике используются иные параметры, связанные с множеством КЛП взаимнообратными функциональными зависимостями.

Рассмотрим следующие параметры авторегрессии, связанные с КЛП:

- нули фильтра ЛП, т.е. полюсы z_{*k} передаточной функции;
- коэффициенты отражения $\{r_i\}$, также известные как коэффициенты частной корреляции (PARCOR);
- линейные спектральные корни (ЛСК), также известные как ЛСЧ.

Из теории устойчивости линейных систем известно, что погрешность представления нулей полинома наименьшим образом влияет на частотные характеристики полинома. Поэтому естественное желание использовать для передачи по КС именно эти рациональные параметры полинома $A_M(z)$:

$$A_M(Z) = 1 + \sum_{k=1}^M a_k z^{-k} = \prod_{k=1}^M (1 - z_{0k} z^{-1}); \quad (2.46)$$

где z_{0k} – нули полинома $A_M(z)$.

С помощью нулей $A_M(z)$ легко контролировать устойчивость фильтра синтеза $H(z)$: согласно условиям устойчивости [106, 166] полюсы z_{*k} передаточной функции $H(z)$, т.е. нули полинома $A_M(z)$, должны находиться внутри круга единичного радиуса на плоскости Z , т.е. для устойчивости необходимо, чтобы выполнялось строгое неравенство:

$$|z_{0k}| = |z_{*k}| = |r_{*k} e^{j\hat{w}_{0k}T}| = r_{*k} < 1, 1 \leq k \leq M; \quad (2.47)$$

где r_{*k} – радиус k -го полюса $H(z)$, т.е. k -го нуля $A_M(z)$, M – порядок передаточной функции $H(z)$.

Расположение нулей $A_M(z)$ внутри круга единичного радиуса означает, что $A(z)$ вида (1.3) и $A_M(z)$ вида (2.46) являются минимально фазовыми [106, 166]. Это важное свойство, которое используется при создании систем обработки РС. Однако в системах обработки РС передача непосредственно z_{*k} не используется, поскольку они являются комплексными и для их вычисления необходимо применять комплексную арифметику, а это приводит к значительным временным затратам, пределы которых ограничены реальным временем работы всей системы. Указанные трудности показывают, что прямое использование КЛП непригодно для передачи информации о состоянии фильтра-предсказателя.

Коэффициенты отражения итерационно рассчитываются по выражениям (2.48), восстановление КЛП – по (2.49).

$$\begin{cases} k_i = a_i^{(i)}; \\ a_j^{(i-1)} = \frac{a_j^{(i)} + a_i^{(i)} a_{i-j}^{(i)}}{1 - k_i^2}; 1 \leq j \leq i-1. \\ a_j = a_j^{(p)}, 1 \leq j \leq p. \end{cases} \quad (2.48)$$

$$\begin{cases} a_i^{(i)} = k_i; \\ a_j^{(i)} = a_j^{(i-1)} - k_i a_{i-j}^{(i-1)}, 1 \leq j \leq i-1, i = 1, 2, \dots, p \\ a_j = a_j^p, 1 \leq j \leq p \end{cases} \quad (2.49)$$

Для коэффициентов отражения характерны следующие особенности:

- чувствительность АЧХ к изменению $\{k_i\}$ существенно ниже, чем к изменению $\{a_k\}$;
- они позволяют легко контролировать устойчивость фильтра на приеме, поскольку условием устойчивости является выполнение неравенства: $|k_i| \leq 1$.
- диапазон их значений не изменяется с ростом порядка ЛП N .

Тем не менее, коэффициенты отражения обычно не передаются. Исследования [109, 186] показали, что равномерное квантование (шаг квантования $Q=\text{const}$) коэффициентов отражения не является наилучшим, так как для минимизации искажений АЧХ фильтра синтеза значения $\{r_i\}$, близкие к единице, должны передаваться с большей точностью, чем значения, близкие к нулю. Отсюда следует необходимость перехода к неравномерному квантованию.

Для устранения этого недостатка используются нелинейные функции – логарифмы отношения (LAR – *Log-Area-Ratio*) (2.50) и арксинусы (IS – *Inverse Sine*) (2.51) коэффициентов отражения:

$$\begin{cases} g_i = \lg \frac{1-k_i}{1+k_i}, 1 < i < M; \\ k_i = \frac{1-10^g}{1+10^g}, 1 < i < M. \end{cases} \quad (2.50)$$

$$\begin{aligned} s_i &= \arcsin(k_i), 1 < i < M; \\ k_i &= \sin(s_i), 1 < i < M. \end{aligned} \quad (2.51)$$

Свойства логарифмов отношений и арксинусов коэффициентов отражения хорошо изучены. Для этих параметров используется равномерное скалярное квантование. Однако они имеют два общих недостатка [109, 110]:

1. Для минимизации искажения АЧХ фильтра синтеза требуется в среднем 4 бита на коэффициент; следовательно, для передачи информации о коэффициентах фильтра 10-го порядка необходимо около 35–40 бит/кадр. Таким образом, на передачу информации о других параметрах остается очень небольшое количество бит.

2. Межкадровая корреляция коэффициентов ЛП (которая присутствует у медленно изменяющихся во времени компонентов речи), не характерна для логарифмов отношения и арксинусов коэффициентов отражения, поэтому трудно предсказать значение параметров по их величине на предыдущем кадре, что приводит к не полному устранению избыточности о РС. Для описания передаточной функции наиболее предпочтительными являются параметры, предложенные Итакурой [109, 110] и названные ЛСЧ или ЛСК [204].

Общий порядок расчета ЛСЧ заключается в следующем. Из полинома передаточной функции образуются новые полиномы $P(z)$ и $Q(z)$ (2.52).

$$\begin{cases} P(Z) = A_M(Z) + Z^{-(M+1)} A_M(Z^{-1}); \\ Q(Z) = A_M(Z) - Z^{-(M+1)} A_M(Z^{-1}); \end{cases} \quad (2.52)$$

Раскрыв скобки, получим (2.53).

$$\begin{cases} P(Z) = 1 + \sum_{k=1}^M a_k Z^{-k} + Z^{-(M+1)} \left[1 + \sum_{k=1}^M a_k Z^k \right] = 1 + \sum_{k=1}^M a_k Z^{-k} + Z^{-(M+1)} + \sum_{k=1}^M a_k Z^{k-(M+1)}; \\ Q(Z) = 1 + \sum_{k=1}^M a_k Z^{-k} - Z^{-(M+1)} \left[1 + \sum_{k=1}^M a_k Z^k \right] = 1 + \sum_{k=1}^M a_k Z^{-k} - Z^{-(M+1)} - \sum_{k=1}^M a_k Z^{k-(M+1)}. \end{cases} \quad (2.53)$$

Перегруппировав члены, входящие в данные выражения получим (2.54).

$$\begin{cases} P(Z) = Z^{-(M+1)} + (a_1 + a_M)Z^{-M} + \dots + (a_1 + a_M)Z^{-1} + 1; \\ Q(Z) = -Z^{-(M+1)} - (a_1 - a_M)Z^{-M} - \dots - (a_1 - a_M)Z^{-1} + 1 \end{cases} \quad (2.54)$$

Для того чтобы вернуться к исходному полиному, определяющему передаточную функцию голосового тракта, необходимо выполнить (2.55).

$$A_M(Z) = \frac{P(Z) + Q(Z)}{2}; \quad (2.55)$$

Полиномы в выражении (2.54) имеют тривиальные вещественные корни: $z_{oP} = -1$ и $z_{oQ} = 1$, для исключения которых нужно разделить $P(z)$ на $1 + z^{-1}$, а $Q(z)$ на $1 - z^{-1}$. Тогда получим выражения (2.56) и (2.57):

$$P'(Z) = \frac{P(Z)}{1 + Z^{-1}} = b_0 Z^{-M} + b_1 Z^{-(M-1)} + \dots + b_{\frac{M}{2}-1} Z^{-\left(\frac{M}{2}+1\right)} + b_{\frac{M}{2}} Z^{-\frac{M}{2}} + b_{\frac{M}{2}-1} Z^{-\left(\frac{M}{2}-1\right)} + \dots + b_1 Z^{-1} + b_0, \quad (2.56)$$

где $b_0 = 1, b_k = a_k + a_{M+1-k} - b_{k-1}$ при $k = 1, \dots, \frac{M}{2}$.

$$Q'(Z) = \frac{Q(Z)}{1 - Z^{-1}} = c_0 Z^{-M} + c_1 Z^{-(M-1)} + \dots + c_{\frac{M}{2}-1} Z^{-\left(\frac{M}{2}+1\right)} + c_{\frac{M}{2}} Z^{-\frac{M}{2}} + c_{\frac{M}{2}-1} Z^{-\left(\frac{M}{2}-1\right)} + \dots + c_1 Z^{-1} + c_0, \quad (2.57)$$

где $c_0 = 1, c_k = a_k + a_{M+1-k} - c_{k-1}$ при $k = 1, \dots, \frac{M}{2}$.

Приведя подобные члены, переходим к выражению (2.58).

$$\begin{cases} P'(z) = Z^{-\frac{M}{2}} \left[b_0 \left(Z^{-\frac{M}{2}} + Z^{\frac{M}{2}} \right) + b_1 \left(Z^{-\left(\frac{M-1}{2}\right)} + Z^{\left(\frac{M-1}{2}\right)} \right) + \dots + b_{\frac{M}{2}} \right]; \\ Q'(z) = Z^{-\frac{M}{2}} \left[c_0 \left(Z^{-\frac{M}{2}} + Z^{\frac{M}{2}} \right) + c_1 \left(Z^{-\left(\frac{M-1}{2}\right)} + Z^{\left(\frac{M-1}{2}\right)} \right) + \dots + c_{\frac{M}{2}} \right] \end{cases} \quad (2.58)$$

Полагая, что $z = e^{j\hat{\omega}}$, $\hat{\omega} = \omega T = 2\pi fT$ и используя соотношение (2.59) получим (2.60):

$$z + z^{-1} = 2 \cos \hat{\omega}; \quad (2.59)$$

$$\begin{cases} P'(\hat{\omega}) = 2e^{-j\frac{M}{2}\hat{\omega}} \left[b_0 \cos\left(\frac{M}{2}\hat{\omega}\right) + b_1 \cos\left(\frac{M-2}{2}\hat{\omega}\right) + \dots + \frac{1}{2}b_{\frac{M}{2}} \right]; \\ Q'(\hat{\omega}) = 2e^{-j\frac{M}{2}\hat{\omega}} \left[c_0 \cos\left(\frac{M}{2}\hat{\omega}\right) + c_1 \cos\left(\frac{M-2}{2}\hat{\omega}\right) + \dots + \frac{1}{2}c_{\frac{M}{2}} \right]; \end{cases} \quad (2.60)$$

Полиномы (2.60) содержат частотно-зависимый множитель $e^{-j\frac{M}{2}\hat{\omega}}$, не влияющий на значения сумм, нули которых определяют корни уравнений, а масштабирующий множитель 2 также не влияет на значения нулей полиномов. Таким образом, при вычислении ЛСК эти множители можно не рассматривать, а при восстановлении КЛП учитывать только множитель 2.

Вводя замену $x = \cos \hat{\omega}$ и применяя формулы кратных дуг [164, 165] можно перейти к алгебраическим полиномам (2.61) и (2.62).

$$P'(x) = \sum_{i=0}^{M/2} d_i x^i; \quad (2.61)$$

$$Q'(x) = \sum_{i=0}^{M/2} q_i x^i; \quad (2.62)$$

Коэффициенты $\{d_i\}$ и $\{q_i\}$ полиномов (2.61) и (2.62) линейно зависят от коэффициентов $\{b_k\}$ и $\{c_k\}$ соответственно. Например, для $M = 10$ они определяются выражениями (2.63) и (2.64).

$$\begin{cases} d_5 = 16b_0; \\ d_4 = 8b_1; \\ d_3 = 4b_2 - 20b_0; \\ d_2 = 2b_3 - 8b_1; \\ d_1 = 5b_0 - 3b_2 + b_4; \\ d_0 = b_1 - b_3 + 0,5b_5. \end{cases} \quad (2.63)$$

$$\begin{cases} q_5 = 16c_0; \\ q_4 = 8c_1; \\ q_3 = 4c_2 - 20c_0; \\ q_2 = 2c_3 - 8c_1; \\ q_1 = 5c_0 - 3c_2 + c_4; \\ c_0 = c_1 - c_3 + 0,5c_5. \end{cases} \quad (2.64)$$

Общее число корней равно M . Итакурой было доказано, что все эти корни различны, вещественны, лежат в интервале $[-1,1]$ и чередуются в этом интервале. В свою очередь это означает, что корни полиномов $P(z)$, $Q(z)$ также различны, вещественны и чередуются. Учитывая, что $x = \cos \hat{\omega}$, по найденным корням нетрудно вычислить искомые частоты выражение (2.65). Корни z_{iP} , z_{iQ} полиномов $P(z)$ и $Q(z)$ имеют вид, представленный в выражении (2.66).

$$\hat{w}_{i|P=0} = \arccos(x_{i|P=0}), \hat{w}_{i|Q=0} = \arccos(x_{i|Q=0}), i = 1, 2, \dots, \frac{M}{2}. \quad (2.65)$$

$$Z_{iP} = e^{j\hat{w}_i P_i}; \quad Z_{iQ} = e^{j\hat{w}_i Q_i}; \quad (2.66)$$

Значения, представленные (2.66), лежат на единичной окружности z -плоскости (их радиусы $r=1$) и соответственно отображаются на частотную ось, что показано на (рисунке 2.8). При этом общее число корней полиномов (2.58) на интервале $[0, \pi]$ равно $(M+2)$, а число тривиальных корней равно 2. Таким образом, общее число нетривиальных корней равно M .

Итакурой были доказаны важнейшие свойства корней полиномов $P(z)$ и $Q(z)$: корни $\hat{w}_{i|P}$ и $\hat{w}_{i|Q}$ полиномов простые и лежат на единичной окружности; корни

полиномов принадлежат интервалу $[0, \pi]$ и чередуются, т.е. за корнем одного полинома следует корень другого [92, 165]. Частоты $\hat{\omega}_{i|P}$ и $\hat{\omega}_{i|Q}$ называют также ЛСП, два соседних корня, принадлежащих разным полиномам, образуют ЛСП. Зная ЛСК, нетрудно восстановить полиномы $P'(z)$ и $Q'(z)$, от которых затем перейти к полиномам $P(z)$ и $Q(z)$ и далее к $A(z)$ [208]. Пример расположения ЛСК на единичной окружности и на частотной оси для $M = 4$, показан на рисунке 2.8, где точками отображены корни $P(z)$, а кружочками – корни полинома $Q(z)$. ЛСК являются вещественными и различными, что не только облегчает их кодирование, но и упрощает контроль над устойчивостью: чередование ЛСК является признаком устойчивости полученного фильтра. На однородных сегментах анализа РС значения ЛСК не претерпевают значительного изменения и не выходят за границы интервалов локализации (рисунок 2.8), что упрощает их поиск и подстройку.

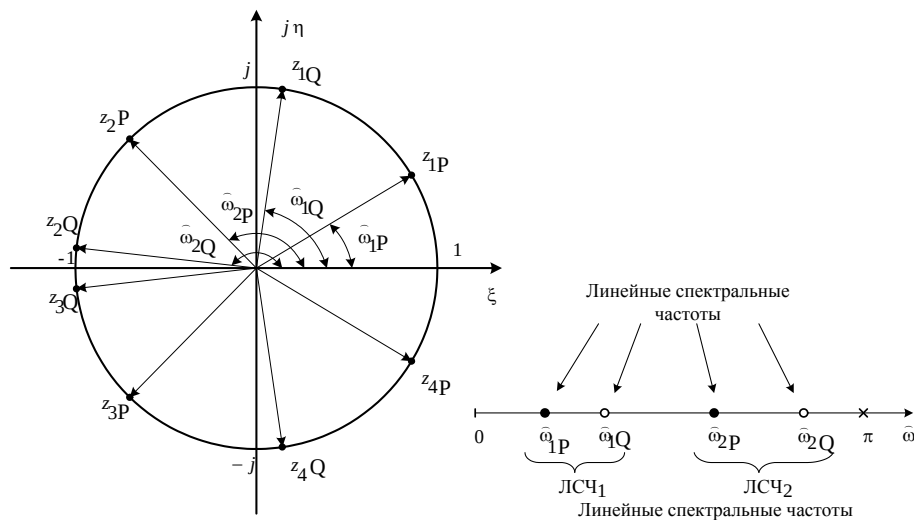


Рисунок 2.8 – Пример расположения нулей полиномов при $M = 4$

При обработке речевых данных одной из основных задач является их качественное и компактное представление для дальнейшего использования при передаче по КС. Решение данных задач позволит увеличить пропускную способность линейных трактов и каналов передачи, улучшить качественные характеристики обрабатываемых РС.

Таким образом, принципиальным при анализе РС на однородных участках РС при использовании ЛСК является формирование из исходного устойчивого полинома $A_M(z)$ нового полинома $A_M(z)$ с монотонной линейной ФЧХ путем умножения первого на z^R , т. е. имеет место преобразование вида:

$$\hat{A}_M(z) = z^R A_M(z) = z^R \sum_{k=0}^M a_k z^{-k} = \sum_{k=0}^M a_k z^{R-k}, \text{ где } a_0 = 1, \frac{M}{2} \leq R \leq M. \quad (2.67)$$

В частотной области полином (2.67) может быть представлен в виде двух составляющих – вещественной и мнимой частями:

$$\hat{A}_M(z) |_{z=e^{j\hat{w}}} = \operatorname{Re}\{\hat{A}_M(z) |_{z=e^{j\hat{w}}}\} + j \operatorname{Im}\{\hat{A}_M(z) |_{z=e^{j\hat{w}}}\}, \quad (2.68)$$

$$\text{где} \quad \begin{cases} \operatorname{Re}\{\hat{A}_M(z) |_{z=e^{j\hat{w}}}\} = P(\hat{w}) = \sum_{k=0}^M a_k \cos[(R-k)\hat{w}] \\ \operatorname{Im}\{\hat{A}_M(z) |_{z=e^{j\hat{w}}}\} = Q(\hat{w}) = \sum_{k=0}^M a_k \sin[(R-k)\hat{w}], \end{cases} \quad (2.69)$$

Основой является утверждение о том, что вещественная часть полинома

$\hat{A}_M(z) |_{z=e^{j\hat{w}}}$ имеет нули на частотах $\hat{w}_{i|P}$, при которых $\arg\{P(\hat{w})\}$ принимает значения $(2k+1)\frac{\pi}{2}$, $k=0,1,\dots,d=\left[R-\frac{1}{2}\right]$, а мнимая часть этого же полинома – на ча-

стотах $\hat{w}_{i|Q}$, при которых $\arg\{Q(\hat{w})\}$ равен $2k\frac{\pi}{2}$, $k=0,1,\dots,d=R$.

Содержание этого утверждения состоит в том, что вещественная $P(\hat{w})$ и мнимая $Q(\hat{w})$ части полинома $z^R A_M(z)$ имеют на интервале $[0, \pi]$ только вещественные корни. Эти корни простые, различные и чередуются [81].

ЛСК могут быть получены из уравнений (2.70):

$$\begin{cases} P(\hat{w}) = \sum_{k=0}^M a_k \cos[(R-k)\hat{w}] \\ Q(\hat{w}) = \sum_{k=0}^M a_k \sin[(R-k)\hat{w}] \end{cases} \quad (2.70)$$

при $\frac{M}{2} \leq R \leq M$, R – целое.

Используя известные соотношения кратных дуг [80], уравнения (2.70) можно привести к виду:

$$\begin{cases} P(\hat{w}) = \sum_{k=0}^R b_k \cos^k [\hat{w}] = 0; \\ Q(\hat{w}) = \sin[\hat{w}] \sum_{k=0}^{R-1} c_k \cos^k [\hat{w}] = 0. \end{cases} \quad (2.71)$$

Коэффициенты, как уже было показано, $\{b_i\}$ и $\{c_i\}$ зависят от КЛП.

Из уравнений (2.71) видно, что вне зависимости от значения константы R мнимый полином $Q(\hat{w})$ имеет два тривиальных нуля при $\hat{w} = 0$ и $\hat{w} = \pi$, которые могут не рассматриваться. С другой стороны, значения $\sin \hat{w}$ при $0 < \hat{w} < \pi$, представляют собой частотно-зависимые множители, не влияющие на значения сумм, нули которых определяют нетривиальные корни уравнения.

Показатель R и количество нетривиальных корней многочленов $P(\hat{w})$ и $Q(\hat{w})$ на интервале $[0, \pi]$ находятся в следующей зависимости: общее число корней полиномов $P(\hat{w})$ и $Q(\hat{w})$ на интервале $[0, \pi]$ равно $2R + 1$. Число тривиальных корней полиномов равно 2. Тогда общее число нетривиальных корней $2R - 1$ (таблица 2.1).

Таблица 2.1 – Распределение ЛСК в зависимости от показателя R

Показатель полинома z^R	Вещественная часть $P(\hat{w})$		Мнимая часть $Q(\hat{w})$	
	число нетрив. корней	число трив. корней	число нетрив. корней	число трив. корней
$R - \text{целое}$	R	—	$R - 1$	2

Таким образом, в зависимости от типа переменной R , для восстановления КЛП $\{a_k\}$ достаточно знать только нетривиальные корни уравнения (2.75) т. е. корни уравнений:

$$\begin{cases} P'(\hat{w}) = \sum_{k=0}^R b_k \cos^k[\hat{w}] = 0; \\ Q'(\hat{w}) = \sum_{k=0}^{R-1} c_k \cos^k[\hat{w}] = 0. \end{cases} \quad (2.72)$$

Исследования [91,110,164,165, 197] показали, что кодирование ЛСК более эффективно по сравнению с другими вариантами описания передаточной функции, а то, что ЛСК находятся в частотной области, позволяет учитывать спектральные особенности РС, что очень важно для его восприятия [92, 109].

Установлено, что каждая ЛСП определяет положение соответствующей форманты, что дает возможность осуществлять их межсегментную интерполяцию на однородных сегментах РС с “гладкими” изменениями спектральной плотности мощности. В низкоскоростных системах обработки РС, повышения качества синтезируемой речи можно добиться, используя в качестве параметров для передачи информацию о состоянии КЛП ЛСК.

На (рисунке 2.9) приведен пример поведения ЛСЧ на однородных сегментах РС [37, 40, 44, 55, 57] для $M=10$, в (таблице 2.2) значения ЛСЧ представлены в радианах. Алгоритм, реализующий выделение однородных сегментов РС представлен в шестом разделе данной работы.

Таблица 2.2 –Значения ЛСЧ (в радианах) на однородных сегментах речевого сигнала

N	1:196	1:246	1:296	1:346
1.	0,252	0,252	0,251	0,251
2.	0,37	0,371	0,371	0,372
3.	0,474	0,474	0,474	0,476
4.	0,618	0,62	0,619	0,618
5.	1	0,998	0,989	0,983
6.	1,664	1,623	1,629	1,607
7.	1,98	1,978	1,977	1,974
8.	2,108	2,108	2,102	2,1
9.	2,544	2,538	2,544	2,537
10.	2,82	2,827	2,822	2,823

Задача нахождения значений ЛСК в общем случае сводится к определению нулей тригонометрических полиномиальных функций (2.72). В современных системах передачи речевых сообщений порядок ЛП может принимать значения от 8

до 20, поэтому в зависимости от показателя R порядок полиномов может изменяться в пределах от 4 до 20. Следовательно, для определения нулей $\hat{w}_{i|P'}$ и $\hat{w}_{i|Q'}$ функций $P'(\hat{w})$ и $Q'(\hat{w})$ с заданной точностью на интервале $[0, \pi]$ необходимо применять численные методы решения уравнений. Указанные нули геометрически соответствуют точкам пересечения кривой с осью частот.

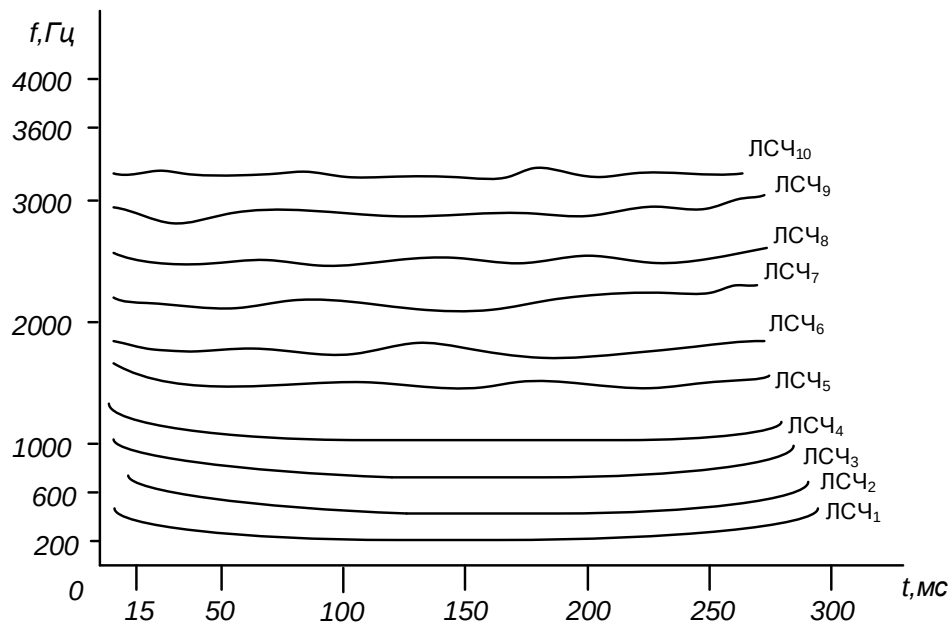


Рисунок 2.9 – Значения линейных спектральных частот на вокализованном однородном сегменте речевого сигнала

В общем случае произвольная полиномиальная функция имеет вид:

$$F(x) = \sum_{i=0}^M a_i x^i = a_M \prod_{i=1}^M (x - x_i), \quad (2.73)$$

где $\{a_i\}$ – коэффициенты функции; $\{x_i\}$ – нули функции; M – порядок функции.

Алгоритм поиска нулей функции (2.77) включает в себя два этапа:

- этап локализации нулей в пределах единого сегмента образования и на под-сегментах, т.е. определение границ таких интервалов, каждый из которых содержит только по одному нулю функции;
- этап вычисления с требуемой точностью нулей функции, принадлежащих найденным интервалам.

Процедура определения границ интервалов, содержащих по одному нулю функции x_i , заключается в поиске таких значений аргумента x_j , в которых исследуемая функция изменяет свой знак только один раз.

Пусть функция $F(x)$ имеет M нулей на интервале $[a, b]$; $\Delta x = x_j - x_{j-1}$ – ширина шага поиска интервалов локализации нулей функции, $i=1, \dots, N$; x_j – j -е значения аргумента функции, $x_j \in [a, b]$, $j = 1, 2, \dots, \frac{b-a}{\Delta x}$.

Тогда правило принятия решения о наличии нуля функции на исследуемом интервале имеет вид:

$$\text{если } F(x_j)F(x_{j-1}) < 0, \quad \text{то } x_{i1} := x_{j-1} \quad \text{и} \quad x_{i2} = x_j. \quad (2.74)$$

Единственность нуля функции на исследуемом интервале (x_{j-1}, x_j) обеспечивается величиной интервала локализации Δx , значение которой должно быть меньше минимального расстояния между соседними нулями функции:

$$\Delta x < \min |x_i - x_{i+1}|. \quad (2.75)$$

Поэтому для минимизации количества контролируемых точек ширина шага поиска интервалов локализации Δx выбирается как минимальное из всех интервалов локализации:

$$\Delta x = \min \{\Delta x\}. \quad (2.76)$$

При одновременном поиске интервалов локализации нулей в пределах единого сегмента образования и на его подсегментах для нескольких функций вида (2.73) эти величины могут быть и не равны, тогда вычисление нулей на интервалах локализации можно выполнить методами дихотомии, линейной интерполяции, касательных или их комбинацией.

При этом необходимо учитывать тот факт, что поиск нулей на интервалах локализации должен проводиться в условиях априорно неизвестного увеличения сегмента анализа однородного участка РС.

Основной составляющей всех численных методов, как на этапе локализации, так и на этапе определения нулей является процедура вычисления значения функции в точке x . Следовательно, вычислительная сложность этих методов определяется: во-первых, сложностью указанной процедуры, т. е. количеством элементарных операций, требуемых для однократного вычисления значения функции, и, во-вторых, количеством обращений к этой процедуре.

Существующие подходы к вычислению ЛСК можно классифицировать по следующим признакам, сгруппированным на (рисунке 2.10).

Рассмотрим модернизацию данных подходов для случая вычисления параметров формирующей модели на однородных сегментах РС [36]. При увеличении границ сегмента анализа в соответствии с таблицей 2.2 происходит вычисление и усреднение значений соответствующих ЛСЧ с целью формирования единого вектора значений [55, 57]. При вычислении достаточно затратным с точки зрения количества вычислительных операций является процедура локализации интервалов нахождения нулей функции.

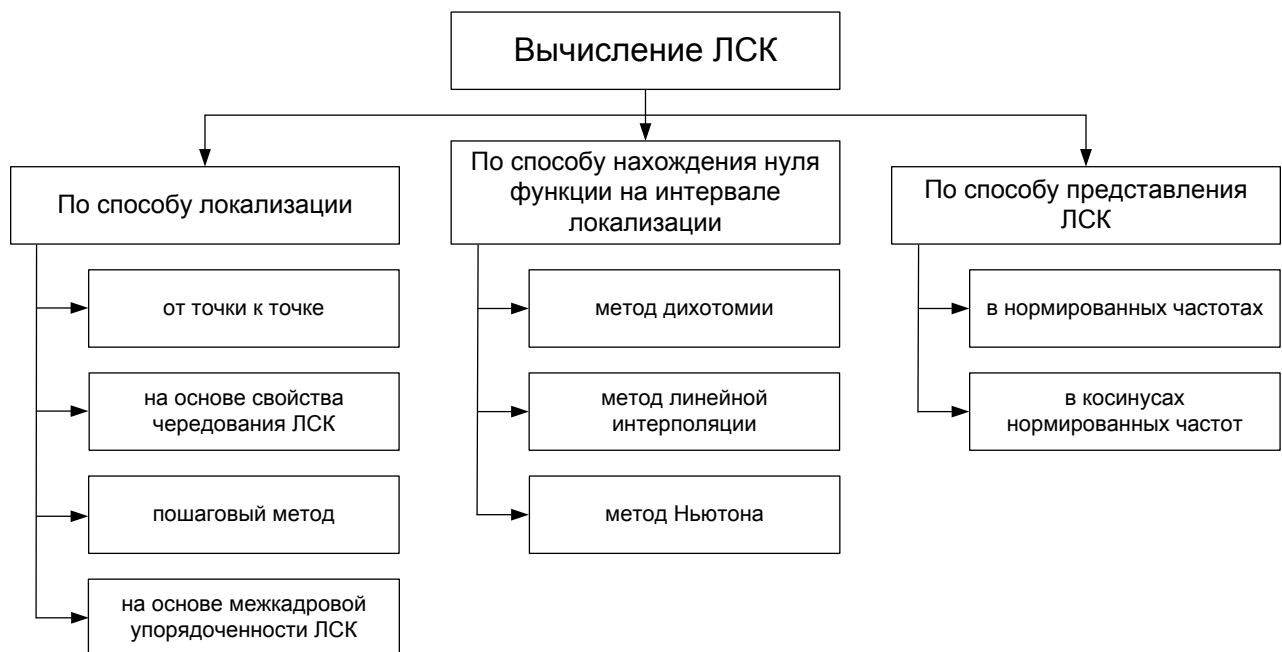


Рисунок 2.10 – Классификация подходов к вычислению ЛСК

Более подробно процесс вычисления нулей $\hat{w}_{i|P'}$ и $\hat{w}_{i|Q'}$ функций $P'(\hat{w})$ и $Q'(\hat{w})$ можно разложить на следующие составляющие:

- определение интервалов локализации нулей $\{\hat{w}_{i|P'}\}$;
- вычисление значений $\{\hat{w}_{i|P'}\}$;
- определение интервалов локализации нулей $\{\hat{w}_{i|Q'}\}$;
- вычисление значений $\{\hat{w}_{i|Q'}\}$.

Локализация нулей функции $P'(\hat{w})$ заключается в проверке с фиксированным шагом поиска диапазона $[0, \pi]$ на изменение знака функции в проверяемых точках. Значение шага поиска не должно превышать меньшего из минимальных расстояний между всеми соседними нулями функции $P'(\hat{w})$:

$$\Delta\hat{w} \leq \min\{\min|\hat{w}_{iP'} - \hat{w}_{i+1P'}|\}, \quad (2.77)$$

где $\hat{w}_{i|P'}$, $\hat{w}_{i+1|P'}$ – i -й и $(i+1)$ -й нули функции, $i = 1, \dots, P_1$; P_1 – количество нулей (порядок) функции $P'(\hat{w})$.

Вычисление нулей функции $P'(\hat{w})$ на найденных интервалах проводят методом дихотомического поиска до достижения заданной точности.

Учитывая, что нули функций $P'(\hat{w})$ и $Q'(\hat{w})$ обладают свойством чередования, границами интервалов локализации $\Delta\hat{w}_{i|Q'}$ нулей $\{\hat{w}_{i|Q'}\}$ являются найденные нули $\{\hat{w}_{i|P'}\}$.

Поэтому вычисление нулей функции $Q'(\hat{w})$ проводятся на интервалах:

$$[\hat{w}_{1P'}, \hat{w}_{2P'},], \dots, [\hat{w}_{iP'}, \hat{w}_{i+1P'},], \dots, [\hat{w}_{P_1, P'}, \pi].$$

Пример поиска интервалов локализации $\Delta\hat{w}_i$ нулей функций $\hat{w}_i$, показан на (рисунке 2.11), где точками отображены нули функций $P'(\hat{w})$ и $Q'(\hat{w})$. Там же для наглядности указаны фиксированные точки $\hat{w}_j$ диапазона $[0, \pi]$ и шаг поиска интервалов локализации $\Delta\hat{w}_i$ функции $P'(\hat{w})$.

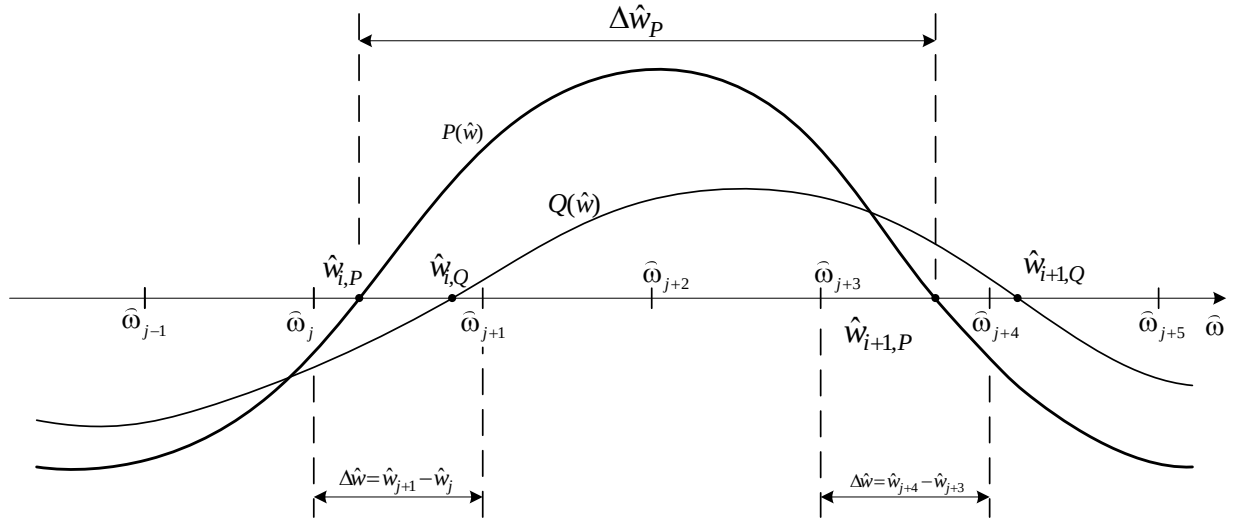


Рисунок 2.11 – Графическое пояснение локализации нулей функций на основе свойства чередования

Если проанализировать выражение (2.76), то можно заметить, что нулями функций $P'(\hat{\omega})$ и $Q'(\hat{\omega})$ фактически являются не значения частот $\hat{\omega}$, а их косинусы. Поэтому, тригонометрические уравнения можно привести к алгебраическому виду, произведя замену $x = \cos \hat{\omega}$. Тогда интервал частотной оси $[0, \pi]$ в области x отобразится в интервал $[-1, 1]$, а выражение (2.76) примет вид (2.78):

$$\begin{cases} P'(x) = \sum_{i=0}^R b_i x^i = 0 \\ Q'(x) = \sum_{i=0}^{R-1} c_i x^i = 0 \end{cases} \quad (2.78)$$

Такое представление снимает необходимость выполнения операции вычисления косинуса частоты при вычислении нулей функции и операции арккосинуса при восстановлении коэффициентов $\{b_i\}$ и $\{c_i\}$, что снижает затрачиваемый вычислительный ресурс.

Таким образом, нули двух полиномиальных функций определяются поочередно за один проход по диапазону $[-1, 1]$. Поиск нулей ведется в сторону уменьшения значения аргумента x . Значение шага поиска интервалов локализации Δx не должно превышать половины наименьшего из минимальных расстояний между соседними нулями одной функции:

$$\Delta x \leq \frac{1}{2} \min[\min|x_{i-1} - x_i|, \min|x_i - x_{i+1}|],$$

где: x_i – нули функции $P'(x)$ или $Q'(x)$, $i = 1, 2, \dots, P$; P – порядок функции;

Количество фиксированных точек на интервале $[-1, 1]$ определяется выражением:

$$Q = \frac{2}{\Delta x_{шаг}} + 1. \quad (2.79)$$

Значения интервалов локализации нулей функций изменяется в пределах $0 < \Delta x_{лок i} < 2\Delta x_{шаг}$.

Алгоритм поиска нулей функций включает в себя следующие шаги:

1. Определение границ интервала, содержащего первый нуль x_i функции $P'(x)$.
2. Вычисление значения нуля на найденном интервале с требуемой точностью одним из известных методов (дихотомии, линейной интерполяции).
3. Поиск интервала локализации первого нуля функции $Q'(x)$, который начинается со значения аргумента, принятого в качестве предыдущего найденного нуля. Если был найден нуль x_i функции $Q'(x)$, то далее определяют интервал локализации Δx второго нуля x_{i+1} $P'(x)$.

Шаги чередования повторяются, пока все x_i нули не будут найдены.

Пример поиска интервалов локализации Δx нулей функций x_i , показан на (рисунке 2.12), где точками отображены нули функций $P'(x)$ и $Q'(x)$. Там же для наглядности указаны фиксированные точки x_j диапазона $[-1, 1]$ и шаг поиска интервалов локализации Δx .

Особенностью такого подхода является то, что при определении интервала локализации $(i+1)$ -го нуля функции $P'(x)$ отправной точкой служит найденное значение i -го нуля функции $Q'(x)$, а при определении интервала локализации $(i+1)$ -го нуля функции $Q'(x)$ отправной точкой служит значение $(i+1)$ -го нуля функции $P'(x)$, положения которых на однородном сегменте от подсегмента к подсегменту изменяются незначительно.

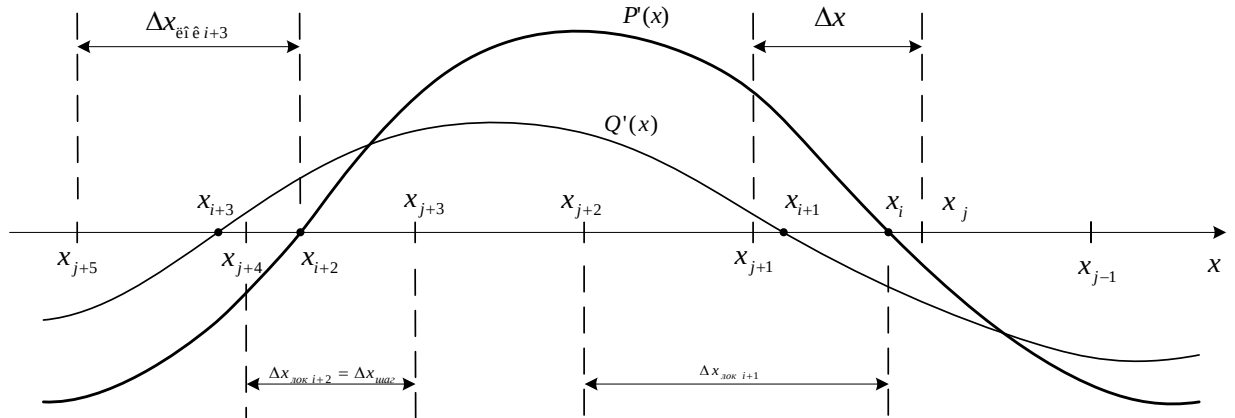


Рисунок 2.12– Графическое пояснение локализации нулей функций пошаговым методом

Кроме того, на этапе локализации максимальное количество обращений к процедуре вычисления значения функции всегда фиксировано и определяется числом проверяемых точек диапазона $[-1,1]$ и числом повторно проверяемых точек, которое определяется числом нулей функций:

$$K_{\text{Л макс}} = T + P_1 + P_2 .$$

где $K_{\text{Л макс}}$ – максимальное количество обращений на этапе локализации;

T – число фиксированных точек; P_1, P_2 – число нулей (порядки) функций.

На этапе уточнения значения нулей число обращений к процедуре вычисления функции зависит от применяемого метода и числа нулей.

$$K_{\text{В макс}} = K_{\text{В}} (P_1 + P_2) .$$

где $K_{\text{В макс}}$ – число обращений на этапе уточнения значений нулей;

$K_{\text{В}}$ – число обращений для определения нуля функций с заданной точностью;

P_1, P_2 – число нулей функций.

Основным достоинством представленного подхода является ограничение на максимум количества обращений к процедуре вычисления функции, который определяется выражением:

$$K_{\text{макс.теор.}} = K_{\text{Л макс}} + K_{\text{В макс}} = \frac{2}{\Delta x_{\text{шаг}}} + 1 + (1 + K_{\epsilon})(P_1 + P_2) .$$

Также необходимо учитывать свойство упорядоченности нулей функций $P'(x)$ и $Q'(x)$, которое не меняется на длительности однородного сегмента РС и на его подсегментах: на исследуемом кадре нули одной функции находятся между значениями нулей другой функции на предыдущем кадре (2.80):

$$x_{i-1}^{(n-1)} < x_i^{(n)} < x_{i+1}^{(n-1)}, \quad (2.80)$$

где $x_i^{(n)}$ – i -й нуль функций на n -ом кадре;

$x_{i-1}^{(n-1)}, x_{i+1}^{(n-1)}$ – $(i-1)$ -й и $(i+1)$ -й нули функций на $(n-1)$ -ом кадре;

$i = 2, \dots, N-1$, N – порядок ЛП.

Согласно [165], в зависимости от порядка ЛП условие межкадровой упорядоченности выполняется в большинстве случаев. Таким образом, на однородных сегментах РС задача локализации нулей функций $P'(x)$ и $Q'(x)$ сводится к проверке условия (2.80).

Контроль выполнения условия упорядоченности может быть проведен путем сравнения знаков функций в точках $\{-1, x_2^{(n-1)}, x_4^{(n-1)}, \dots, x_{N-2}^{(n-1)}, x_N^{(n-1)}, 1\}$ при вычислении нулей функции $P'(x)$ $\{x_1^{(n)}, x_3^{(n)}, \dots, x_{N-3}^{(n)}, x_{N-1}^{(n)}\}$, или $\{-1, x_1^{(n-1)}, x_3^{(n-1)}, \dots, x_{N-3}^{(n-1)}, x_{N-1}^{(n-1)}, 1\}$ при вычислении нулей функции $Q'(x)$ $\{x_1^{(n)}, x_3^{(n)}, \dots, x_{N-3}^{(n)}, x_{N-1}^{(n)}\}$.

Анализ существующих методов в данной предметной области показывает, что основная часть вычислительного ресурса приходится на многократные вычисления значений полиномиальных функции, причем большая его составляющая связана с определением интервалов локализации нулей. Кроме того, поиск интервалов локализации построен таким образом, что отправной точкой для поиска искомого нуля является значение предыдущего найденного нуля. Это приводит к тому, что в большинстве случаев результат проверки интервала на изменение знака функции является отрицательным и фактически приходится проверять весь диапазон возможных значений нулей, что значительно увеличивает вычислительную сложность

алгоритма, уменьшая при этом точность вычисления соответствующего нуля. При нахождении нулей функций $P'(x)$ и $Q'(x)$ не учитывается их поведение (знак) на интервалах и не накапливается статистика распределения нулей функций, дополнительные сведения о которых могли бы способствовать уменьшению количества проверяемых интервалов при локализации нулей и позволили бы точнее определять значение ЛСК на однородных сегментах РС при их вычислении.

При вычислении ЛСК процедура поиска нулей функций $P'(x)$ и $Q'(x)$ связана с последовательным решением двух задач:

- во-первых, установления интервалов локализации нулей;
- во-вторых, вычисления с заданной точностью нуля, принадлежащего найденному интервалу.

При вычислении ЛСК необходимо использовать информацию о знаке функции на интервалах между её нулями. Такая информация может способствовать уменьшению количества контролируемых точек при определении интервалов локализации нулей функции [40].

Таким образом:

- в общем случае функции имеют вид (2.81):

$$P'(w) = \sum_{k=0}^P g_k \cos^k \hat{w}, \quad (2.81)$$

где $\{g_k\}$ – коэффициенты функции, P – порядок функции.

- коэффициенты функций вещественные;
- нули функций также вещественны, различны, обладают свойством чередования и лежат на частотной оси в интервале $(0, \pi)$;
- функции не имеют нулей в точках $\hat{w} = 0$ и $\hat{w} = \pi$.

Следовательно, функция вида (2.81) с фиксированными коэффициентами на длительности кадра, сохраняет свой знак в пределах интервалов:

$$(0, w_1), (\hat{w}_1, \hat{w}_2), \dots, (\hat{w}_{P-1}, \hat{w}_P), (\hat{w}_P, \pi) \quad (2.82)$$

где $\{\hat{w}_i\}$ – нули функции, $i = 1, \dots, P$.

Следовательно, знак функции на интервале $[0, \hat{\omega}_1]$ определяет порядок изменения знака функции на всех интервалах между её нулями.

Следовательно, для функции (2.81) возможны два варианта:

- функция положительна на интервале $[0, \hat{\omega}_1)$;
- функция отрицательна на интервале $[0, \hat{\omega}_1)$.

Таким образом, существует неопределенность в поведении функции на интервалах между ее нулями. На (рисунке 2.13) показаны два возможных варианта поведения функции $P'(w)$ пятого порядка.

Для устранения неопределенности в поведении функций $P'(w)$ и $Q'(w)$ на интервалах между их нулями докажем два утверждения.

Утверждение. В точке $\hat{\omega} = 0$ тригонометрические полиномы $P'(\hat{\omega})$ и $Q'(\hat{\omega})$ принимают только положительные значения независимо от значений коэффициентов $\{a_k\}$ и показателя R .

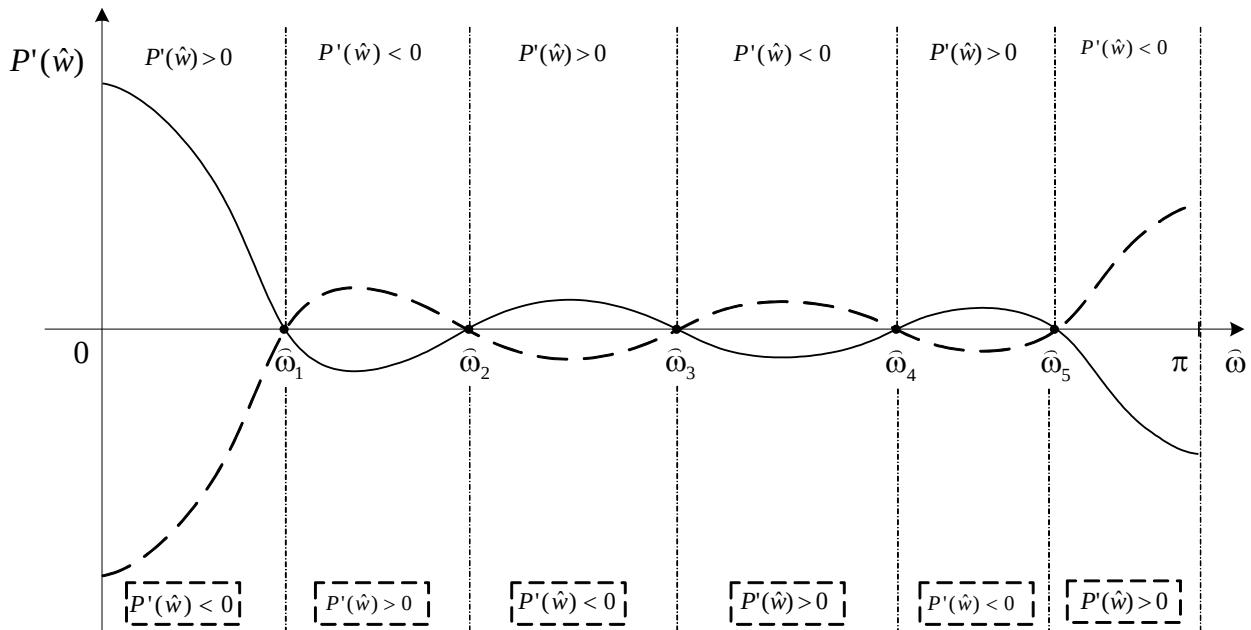


Рисунок 2.13 – Возможные варианты поведения функции пятого порядка на интервалах между нулями

Доказательство. Рассмотрим взаимосвязь вещественной части $\text{Re}\left\{\hat{A}_M(z)\Big|_{z=e^{j\hat{\omega}}}\right\} = P(\hat{\omega})$ полинома и тригонометрического полинома $P'(\hat{\omega})$:

$$P(\hat{\omega}) = \sum_{k=0}^M a_k \cos(R-k)\hat{\omega} = \sum_{k=0}^R \tilde{a}_k \cos^k \hat{\omega} = P'(\hat{\omega}). \quad (2.82)$$

При $\hat{\omega} = 0$ выражение (2.71) принимает вид:

$$P(\hat{\omega})|_{\hat{\omega}=0} = \sum_{k=0}^M a_k = \sum_{k=0}^R \tilde{a}_k = P'(\hat{\omega})|_{\hat{\omega}=0}. \quad (2.83)$$

Из выражений (2.83) следует, что для любых M знак тригонометрической функции $P'(\hat{\omega})$ в точке $\hat{\omega} = 0$ определяется суммой КЛП:

$$P'(\hat{\omega})|_{\hat{\omega}=0} = 1 + \sum_{k=1}^M a_k. \quad (2.84)$$

Сумма КЛП всегда удовлетворяет (2.83), поэтому справедливо неравенство:

$$0 < P'(\hat{\omega})|_{\hat{\omega}=0} < 2^M. \quad (2.85)$$

Полином $P'(\hat{\omega})$ в точке $\hat{\omega} = 0$ принимает только положительные значения независимо от состояния КЛП и показателя R .

В точке $\hat{\omega} = 0$ тригонометрический полином $Q'(\hat{\omega})$ также принимает только положительные значения независимо от значений коэффициентов $\{a_k\}$ и показателя R . Далее рассмотрим взаимосвязь мнимой части $\text{Im}\left\{\hat{A}_M(z)\Big|_{z=e^{j\hat{\omega}}}\right\} = Q(\hat{\omega})$ полинома (2.76) и тригонометрического полинома $Q'(\hat{\omega})$:

$$Q'(\hat{\omega}) = \sum_{k=0}^{R-1} \tilde{c}_k \cos^k \hat{\omega}. \quad (2.86)$$

При $\hat{\omega} = 0$ получим

$$Q'(\hat{\omega}) = \sum_{k=0}^{R-1} \tilde{c}_k. \quad (2.87)$$

Коэффициенты $\{\tilde{c}_k\}$ и $\{\tilde{d}_k\}$ линейно зависят от коэффициентов $\{a_k\}$, но по выражениям (2.87) нельзя однозначно определить знак тригонометрической функции $Q'(\hat{\omega})$ в точке $\hat{\omega} = 0$. Поэтому возможно получение эквивалентного разложения $Q'(\hat{\omega})$, которое содержит КЛП в явном виде для любых R .

Для получения такого разложения проанализируем исходное выражение мнимой части $\text{Im}\left\{\hat{A}_M(z)\Big|_{z=e^{j\hat{\omega}}}\right\} = Q(\hat{\omega})$ полинома (2.75), содержащего КЛП (2.88):

$$Q(\hat{\omega}) = \sum_{k=0}^M a_k \sin[(R-k)\hat{\omega}]. \quad (2.88)$$

Выражение (2.88) в тождественном виде будет иметь следующий вид:

$$Q(\hat{\omega}) = \frac{\sin \hat{\omega}}{\sin \hat{\omega}} \sum_{k=0}^M a_k \sin[(R-k)\hat{\omega}],$$

следовательно

$$Q(\hat{\omega}) = \sin \hat{\omega} \sum_{k=0}^M \left[a_k \frac{\sin(R-k)\hat{\omega}}{\sin \hat{\omega}} \right].$$

Полином $Q(\hat{\omega})$ может быть представлен следующим выражением (2.89):

$$Q(\hat{\omega}) = \sin \hat{\omega} \sum_{k=0}^M \left[a_k \frac{\sin(R-k)\hat{\omega}}{\sin \hat{\omega}} \right] = \sin \hat{\omega} \cdot Q'(\hat{\omega}); \quad (2.89)$$

Далее рассмотрим соотношение (2.90):

$$Q(\hat{\omega}) = \sin \hat{\omega} \sum_{k=0}^{R-1} \tilde{c}_k \cos^k \hat{\omega} = \sin \hat{\omega} \cdot Q'(\hat{\omega}) \quad (2.90)$$

Сравнивая выражения (2.89) и (2.90), получим выражение (2.91)

$$Q'(\hat{\omega}) = \sum_{k=0}^{R-1} \tilde{c}_k \cos^k \hat{\omega} = \sum_{k=0}^M \left[a_k \frac{\sin(R-k)\hat{\omega}}{\sin \hat{\omega}} \right]. \quad (2.91)$$

В соответствии с правилом Лопиталя, имеем:

$$\left. \frac{\sin(R-k)\hat{\omega}}{\sin \hat{\omega}} \right|_{\hat{\omega}=0} = \lim_{\hat{\omega} \rightarrow 0} \frac{(\sin(R-k)\hat{\omega})'}{(\sin \hat{\omega})'} = \lim_{\hat{\omega} \rightarrow 0} \frac{\cos[(R-k)\hat{\omega}] \cdot (R-k)}{\cos \hat{\omega}} = (R-k),$$

поэтому выражение (2.91) при $\hat{\omega} = 0$ принимает вид:

$$Q(\hat{\omega})|_{\hat{\omega}=0} = \sum_{k=0}^{R-1} \tilde{c}_k = \sum_{k=0}^M a_k (R-k). \quad (2.92)$$

В общем случае, знак тригонометрической функции $Q'(\hat{\omega})$ в точке $\hat{\omega} = 0$ будет определяться знаком суммы

$$\sum_{k=0}^M a_k (R-k) = R \sum_{k=0}^M a_k \left(1 - \frac{k}{R} \right).$$

Множитель R не влияет на знак результата, поэтому достаточно определить только знак суммы

$$\sum_{k=0}^M a_k \left(1 - \frac{k}{R}\right). \quad (2.93)$$

Выражение (2.93) представляет собой сумму коэффициентов минимально-фазового полинома вида (2.77), поэтому (2.93) принимает только положительные значения при любых $\frac{M}{2} \leq R \leq M$. Поэтому с учетом (2.77) и (2.91) для (2.92) справедливы следующие условия:

$$0 < Q'(\hat{\omega})|_{\hat{\omega}=0} < 2^M R$$

Следовательно, полином $Q'(\hat{\omega})$ в точке $\hat{\omega} = 0$ принимает только положительные значения независимо от КЛП и показателя R , что и требовалось доказать.

На основании данного утверждения можно сформулировать общее свойство функций $P'(\hat{\omega})$ и $Q'(\hat{\omega})$:

– если нули функции пронумеровать в направлении увеличения частоты, то с повышением ее порядкового номера на однородных сегментах РС функция изменяет знак с положительного на отрицательный при прохождении через нечетные нули; и с отрицательного на положительный – при прохождении через четные нули.

Иллюстрация данного свойства представлена на (рисунке 2.14) для функции пятого порядка.

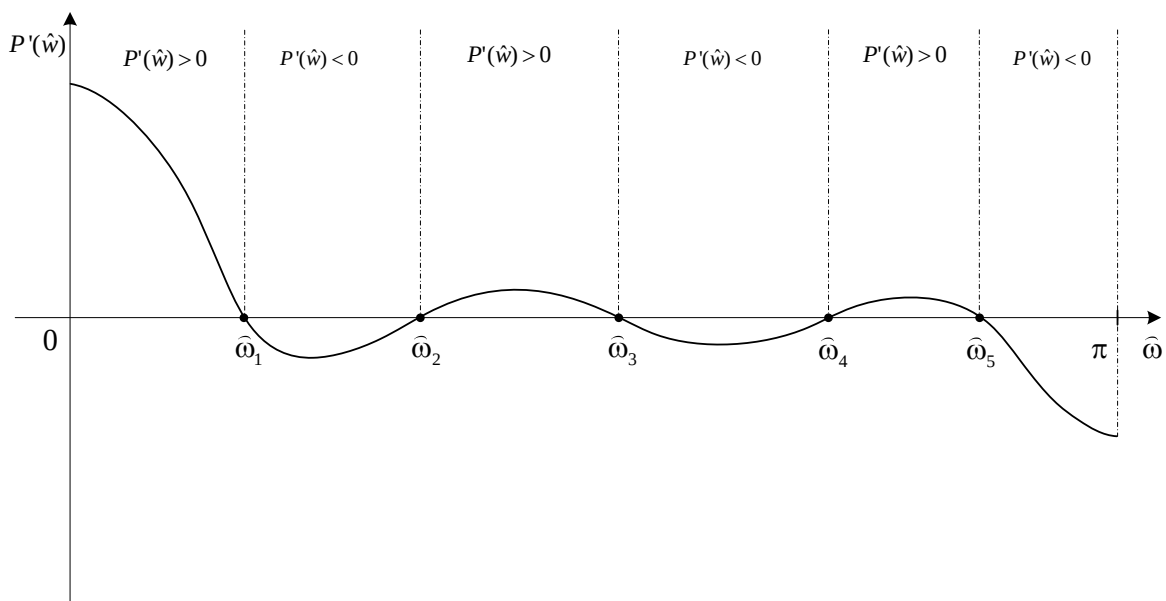


Рисунок 2.14 – Поведение функции $P'(\hat{\omega})$ на интервалах между её нулями

2.5. Оценка и коррекция линейных спектральных частот на однородных сегментах анализа при обработке речевого сигнала

Методы определения интервалов, содержащих единственный нуль функции, используют в качестве исходной точки или значение искомого нуля функции на предыдущем сегменте, или значение предыдущего найденного нуля другой функции на текущем. Однако порядок чередования знаков функций $P'(\hat{\omega})$ и $Q'(\hat{\omega})$ не учитывается ни одним из известных подходов.

Полиномиальные функции $P'(\hat{\omega})$ и $Q'(\hat{\omega})$ в точке $\hat{\omega} = 0$ всегда имеют положительный знак для любых $\{a_k\}$ и $\frac{M}{2} \leq R \leq M$, что определяет порядок изменения знаков функций на интервалах между её нулями.

На основе этого свойства можно утверждать, что по знаку функции с фиксированными коэффициентами, что справедливо на длительности сегмента однородности в некоторой точке $\hat{w}_{oni}$ принадлежащей одному из интервалов

$$\begin{cases} \hat{w}_{oni} \in (\hat{w}_{i-1}, \hat{w}_i) \\ \hat{w}_{oni} \in (\hat{w}_i, \hat{w}_{i+1}) \end{cases}, \quad (2.94)$$

можно определить направление дальнейшего поиска интервала локализации i -го нуля функции. При этом точку, удовлетворяющую условию (2.94), называют опорной. Число контролируемых точек уменьшается, если учитывать направление относительно выбранной опорной точки на подсегменте единого сегмента анализа, соответствующей искомой ЛСЧ. При этом необходимо двигаться для определения интервала локализации соответствующего нуля функции.

В соответствии с предлагаемым подходом для определения направления необходимо вычислить значение полинома в опорной точке, соответствующей искомому нулю функции. В зависимости от знака функции и четности порядкового

номера искомого i -го нуля функции на подсегменте единого сегмента анализа необходимо определить направление дальнейшего поиска интервала локализации по следующему правилу:

- если искомый нуль функции имеет четный порядковый номер и значение функции в опорной точке меньше нуля, то поиск интервала локализации необходимо вести в направлении увеличения значения частоты (рисунок 2.15,а);
- если искомый нуль функции имеет четный порядковый номер и значение функции в опорной точке больше нуля, то поиск интервала локализации необходимо вести в направлении уменьшения значения частоты (рисунок 2.15, б);
- если искомый нуль функции имеет нечетный порядковый номер и значение функции в опорной точке больше нуля, то поиск интервала локализации необходимо вести в направлении увеличения значения частоты (рисунок 2.15, в);

если искомый нуль функции имеет нечетный порядковый номер и значение функции в опорной точке меньше нуля, то поиск интервала локализации нужно вести в направлении уменьшения значения частоты (рисунок 2.15, г).

Определение направлений поиска интервалов локализации для остальных нулей функции проводится аналогично с учетом того, что для каждого из них определение направления должно осуществляться по знаку функции в соответствующих им опорных точках.

Предлагаемый подход к определению направления поиска интервала локализации подразумевает наличие информации об опорных точках. Значение опорной точки принадлежит одному из множеств $\{\hat{\omega}_{i-1}, \hat{\omega}_i\}$ или $\{\hat{\omega}_i, \hat{\omega}_{i+1}\}$. При выборе опорной точки следует учитывать, что на разных сегментах значения $\{\hat{\omega}_i\}$ различны. Вместе с тем, на практике желательно заранее знать значение опорной точки, причем оно должно быть фиксированным для всех сегментов. В качестве таких точек можно использовать значение предыдущего найденного нуля на сегменте. В устройствах обработки и передачи РС порядок полиномов ЛП фиксирован, тогда фиксированы как порядки функций $P'(\hat{\omega})$ и $Q'(\hat{\omega})$, так и количество их нулей. В

процессе анализа изменяются только коэффициенты полинома ЛП и, соответственно, коэффициенты функций $P'(\hat{w})$ и $Q'(\hat{w})$.

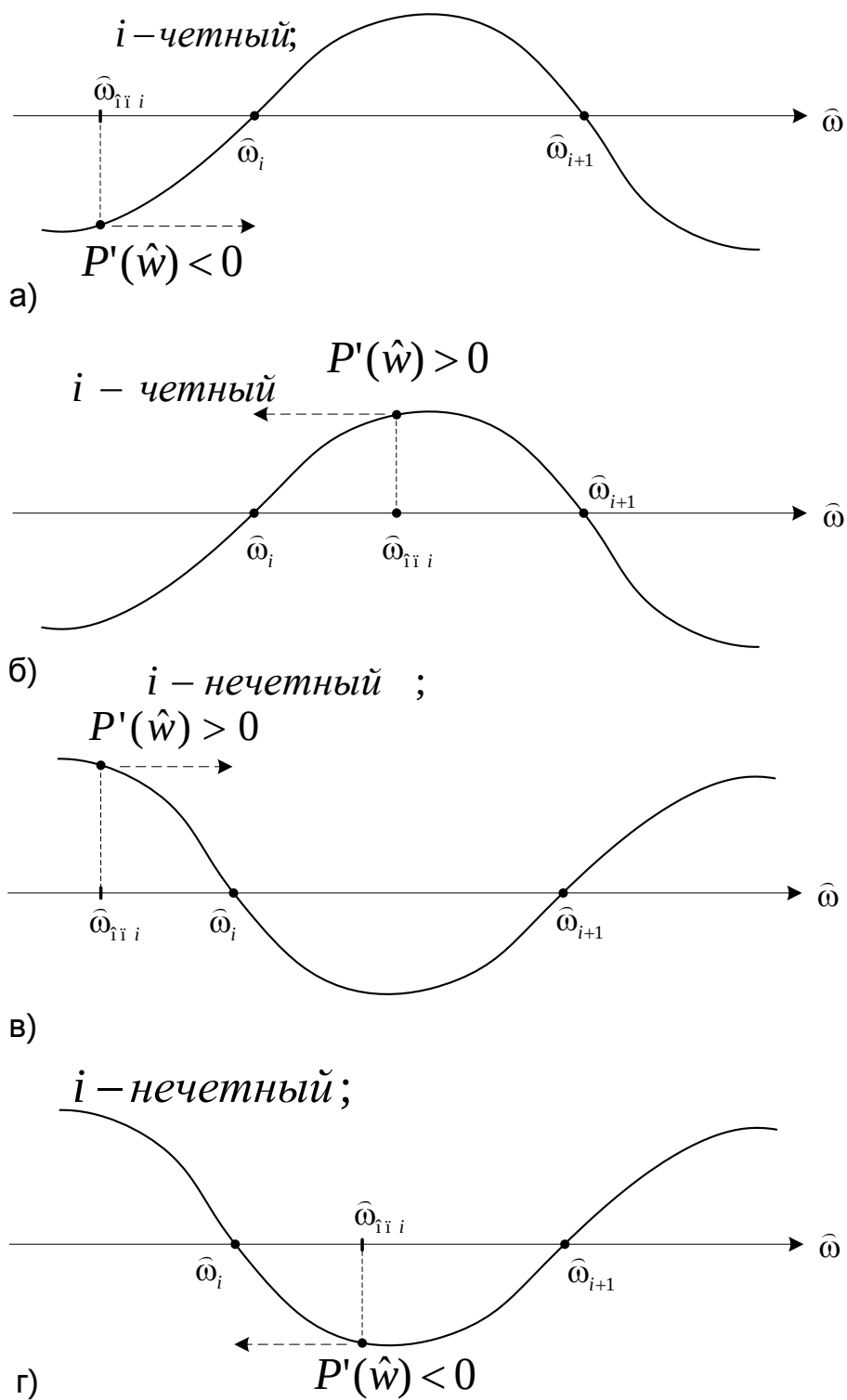


Рисунок 2.15 – Графическое пояснение процедуры выбора направления поиска нулей функций

Учитывая, что значения нулей функций принадлежат множеству $(0, \pi)$, предлагается в качестве опорных точек использовать статистические параметры нулей функций, найденные средневзвешенным значением соответствующих нулей на подсегментах единого однородного сегмента РС.

Теоретически такими параметрами могут быть значения математических ожиданий или медиан распределений нулей функций $P'(\hat{w})$ и $Q'(\hat{w})$ на подсегментах единого однородного сегмента РС.

Таким образом, данный метод позволяет определить интервал локализации соответствующего нуля полиномиальной функции на однородном сегменте РС. В качестве примера рассмотрим применение предлагаемого подхода к вычислению нулей функции $P'(\hat{w})$:

1. Вычисление значения функции в i -ой опорной точке. Значение функции в точке определяется выражением:

$$P'(\hat{w}_{on}) = \sum_{k=0}^{M/2} g_k \cos^k \hat{w}_{on},$$

2. Определение направления поиска i -го нуля функции. Выбор направления дальнейшего поиска осуществляется по следующему правилу:

если $P'(\hat{w}_{on}) > 0$ и i – нечетное, то $\hat{w}_e = \hat{w}_{ion} + \Delta\hat{w}_{шаг}$, $\hat{w}_н = \hat{w}_{ion}$, $m = 1$;

если $P'(\hat{w}_{on}) < 0$ и i – нечетное, то $\hat{w}_e = \hat{w}_{ion}$, $\hat{w}_н = \hat{w}_{ion} - \Delta\hat{w}_{шаг}$, $m = -1$;

если $P'(\hat{w}_{on}) > 0$ и i – четное, то $\hat{w}_e = \hat{w}_{ion}$, $\hat{w}_н = \hat{w}_{ion} - \Delta\hat{w}_{шаг}$, $m = -1$;

если $P'(\hat{w}_{on}) < 0$ и i – четное, то $\hat{w}_e = \hat{w}_{ion} + \Delta\hat{w}_{шаг}$, $\hat{w}_н = \hat{w}_{ion}$, $m = 1$;

где m – идентификатор направления поиска, M – порядок полиномиальной функции, $\{g_k\}$ – коэффициенты полиномиальной функции, $\{\hat{w}_{ion}\}$ – значения опорных точек, $\Delta\hat{w}_{шаг}$ – значение шага локализации, $\{\hat{w}_e\}$ – верхняя граница интервала локализации, $\{\hat{w}_н\}$ – нижняя граница интервала локализации.

3. Вычисление значения функции в краевых точках шага поиска. Значение функции в точках определяется выражениями:

$$\begin{cases} F(\hat{w}_e) = \sum_{k=0}^{M/2} g_k \cos^k \hat{w}_e \\ F(\hat{w}_n) = \sum_{k=0}^{M/2} g_k \cos^k \hat{w}_n \end{cases}$$

4. Определение наличия нуля функции на шаге поиска осуществляется по правилу:

– если $P'(\hat{w}_e)P'(\hat{w}_n) > 0$, то текущий интервал не содержит нуля функции и следует перейти к пункту 5.

– если $P'(\hat{w}_e)P'(\hat{w}_n) < 0$, то текущий интервал содержит нуль функции и следует перейти к пункту 6;

5. Определение новых границ поиска нуля функции осуществляется по правилу:

– если $m = -1$, то $\hat{w}_e = \hat{w}_n$, $\hat{w}_n = \hat{w}_n - \Delta\hat{w}_{шаг}$;

– если $m = 1$, то $\hat{w}_n = \hat{w}_e$, $\hat{w}_e = \hat{w}_e - \Delta\hat{w}_{шаг}$;

После переопределения границ перейти к пункту 3.

6. Определение границ интервала локализации i -го нуля функции:

$$\begin{cases} \hat{w}_e = \hat{w}_e \\ \hat{w}_n = \hat{w}_n \end{cases}$$

7. Определение математических ожиданий распределения интервалов нахождения нулей функций $P'(\hat{w})$ и $Q'(\hat{w})$ на подсегментах единого однородного сегмента РС.

При вычислении значений ЛСК функций $P'(\hat{w})$ и $Q'(\hat{w})$ значение последующего спектрального корня, одного полинома всегда больше, чем значение предыдущего корня другого полинома. При одновременном вычислении ЛСК функций необходимо учитывать этот факт и в качестве опорной точки для нахождения корня $P'(\hat{w})$ целесообразно выбирать значение предыдущего найденного корня соответствующего полинома $Q'(\hat{w})$.

Предлагаемый метод позволяет определить значения ЛСК $P'(\hat{w})$ на однородных сегментах РС, используя их значения на подсегментах.

Предлагаемый подход заключается в следующем:

1) Начальная установка коэффициентов функции: $\{g_k\} = \{b_i\}$ и инициализация значений опорных точек: $\{\hat{w}_{\text{оп}}\} = \text{sort}[\{\hat{w}_{P_1 \text{оп}}\}, \{\hat{w}_{P_2 \text{оп}}\}]$,

где $\text{sort}[*]$ – операция сортировки по убыванию.

2) Инициализация счетчика общего количества нулей $i = 1$.

3) Определение значения текущей проверяемой точки на подсегментах $\hat{w}_B = \hat{w}_{i \text{оп}}$. Если $i = 1$, то перейти к п. 5 алгоритма.

4) Проверка корректности выбора значения текущей точки.

Если значение текущей точки больше значения предыдущего найденного нуля $\hat{w}_B > \hat{w}_{i-1}$, то в качестве текущей точки принимается значение предыдущего нуля $\hat{w}_B = \hat{w}_{i-1}$.

5) Вычисление значения функции в текущей точке по выражению:

$$P(\hat{w}_B) = \sum_{k=0}^{M/2} g_k \hat{w}_B^k.$$

6) Определение направления поиска i -го нуля функции, выбор которого осуществляется по следующему правилу:

если $F(\hat{w}_B) > 0$ и i – нечетное, $\hat{w}_H = \hat{w}_B - \Delta \hat{w}_{\text{шаг}}$, $m = -1$;

если $F(\hat{w}_B) > 0$ и i – четное, то $\hat{w} = \hat{w}_B$, $\hat{w}_B = \hat{w} + \Delta \hat{w}_{\text{шаг}}$, $\hat{w}_H = \hat{w}$, $m = 1$;

если $F(\hat{w}_B) < 0$ и i – нечетное, то $\hat{w} = \hat{w}_B$, $\hat{w}_B = \hat{w} + \Delta \hat{w}_{\text{шаг}}$, $\hat{w}_H = \hat{w}$, $m = 1$;

если $F(\hat{w}_B) < 0$ и i – четное, $\hat{w}_H = \hat{w}_B - \Delta \hat{w}_{\text{шаг}}$, $m = -1$;

где m – идентификатор направления поиска.

7) Вычисление значения функции в краевых точках шага поиска. Значение функции в точках определяется выражениями:

$$P(\hat{w}_B) = \sum_{k=0}^M g_k \hat{w}_B^k, \quad P(\hat{w}_H) = \sum_{k=0}^M g_k \hat{w}_H^k.$$

8) Определение наличия нуля функции на шаге поиска осуществляется по правилу:

– если $P(\hat{w}_B)P(\hat{w}_H) > 0$, то текущий интервал не содержит нуль функции и следует перейти к п. 12 алгоритма.

– если $P(\hat{w}_B)P(\hat{w}_H) < 0$, то текущий интервал содержит нуль функции и следует перейти к п. 10 алгоритма;

9) Определение новых границ поиска нуля функции по правилу:

– если $m = -1$, то $\hat{w}_B = \hat{w}_H$, $\hat{w}_H = \hat{w}_H - \Delta\hat{w}_{\text{шаг}}$;

– если $m = 1$, то $\hat{w}_H = \hat{w}_B$, $\hat{w}_B = \hat{w}_B + \Delta\hat{w}_{\text{шаг}}$;

После переопределения границ перейти к п. 7 алгоритма, иначе к п. 10.

10) Сужение найденного интервала локализации i -го нуля функции методом дихотомии.

11) Определение нуля функции на конечном интервале за один шаг метода линейной интерполяции:

$$\hat{w}_i = \hat{w}_H - \frac{(\hat{w}_B - \hat{w}_H) \cdot P(\hat{w}_H)}{P(\hat{w}_B) - P(\hat{w}_H)}.$$

12) Проверка условия нахождения всех корней функции: если все нули найдены, то завершить работу.

13) Определение математических ожиданий значений нулей функций $P'(\hat{w})$ на подсегментах единого однородного сегмента РС.

14) Оценка нулей $P'(\hat{w})$ в рамках единого однородного сегмента РС.

Следует учесть, что одновременно все нули функций $P'(\hat{w})$ и $Q'(\hat{w})$ не могут принимать одновременно или минимальные или максимальные значения.

На (рисунке 2.16) и (рисунке 2.17) представлены карты нулей и полюсов фильтров синтеза с передаточной функцией вида (1.4), знаменателями, которых яв-

ляются полиномы (1.3). Из рисунков видно, что восстановленные фильтры являются устойчивыми, так как их полюсы расположены внутри круга единичного радиуса z -плоскости.

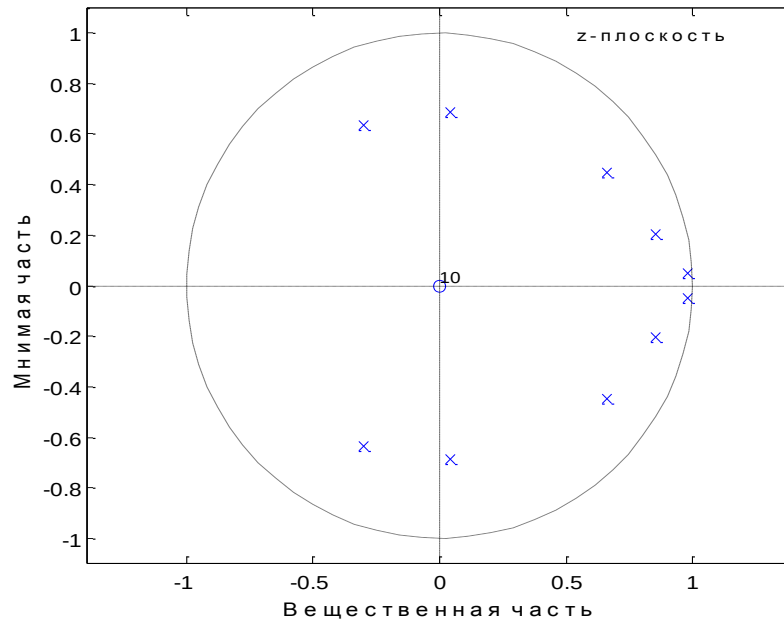


Рисунок 2.16 – Карта нулей и полюсов фильтра синтеза при условии, что нули функций $P'(\hat{w})$ и $Q'(\hat{w})$ принимают минимальные значения

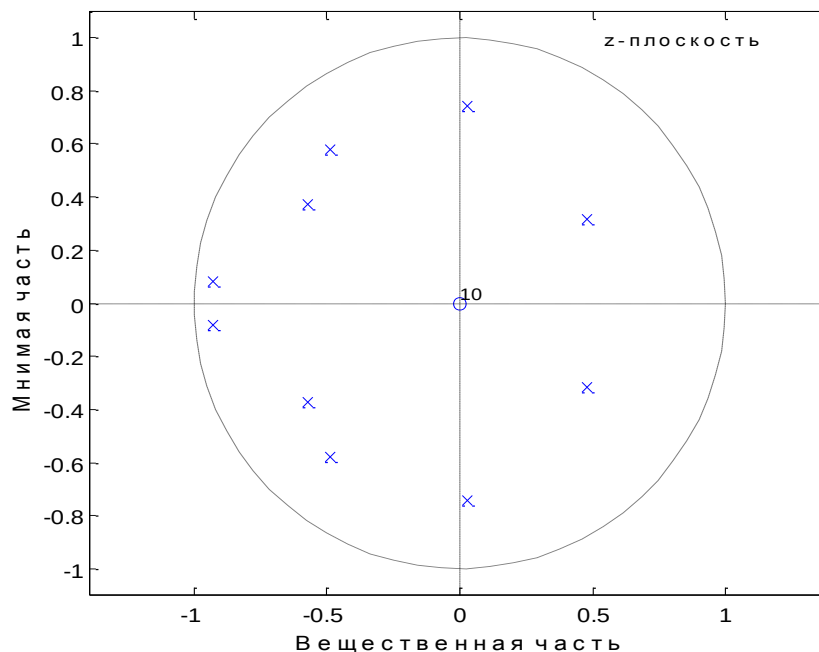


Рисунок 2.17– Карта нулей и полюсов фильтра синтеза при условии, что нули функций $P'(\hat{w})$ и $Q'(\hat{w})$ принимают максимальные значения

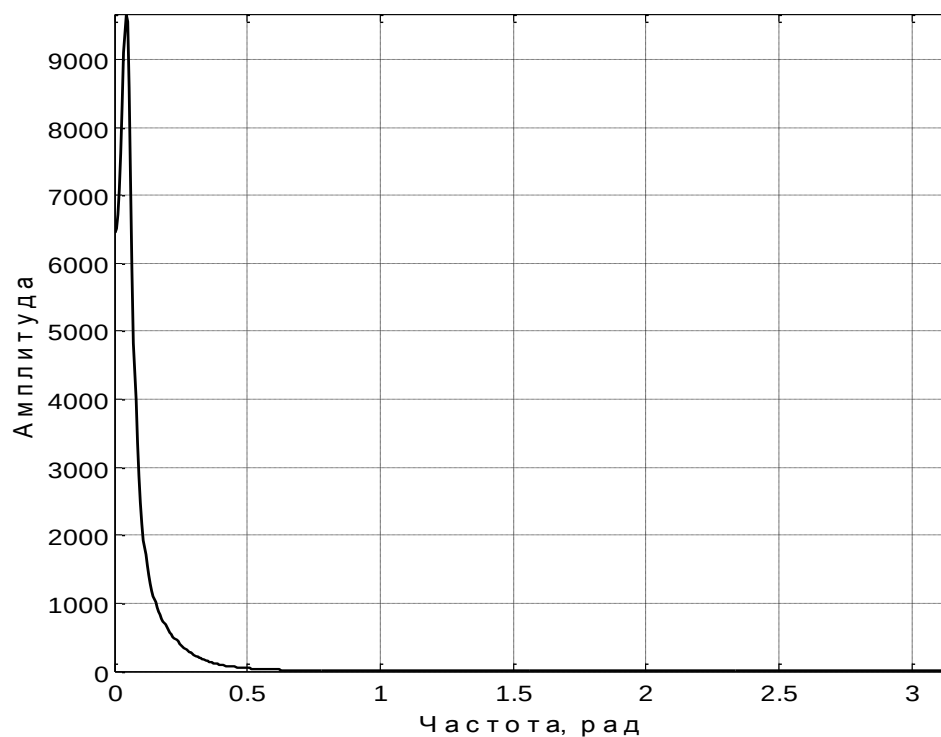


Рисунок 2.18 – АЧХ фильтра синтеза при условии, что нули функций $P'(\hat{\omega})$ и $Q'(\hat{\omega})$ принимают минимальные значения

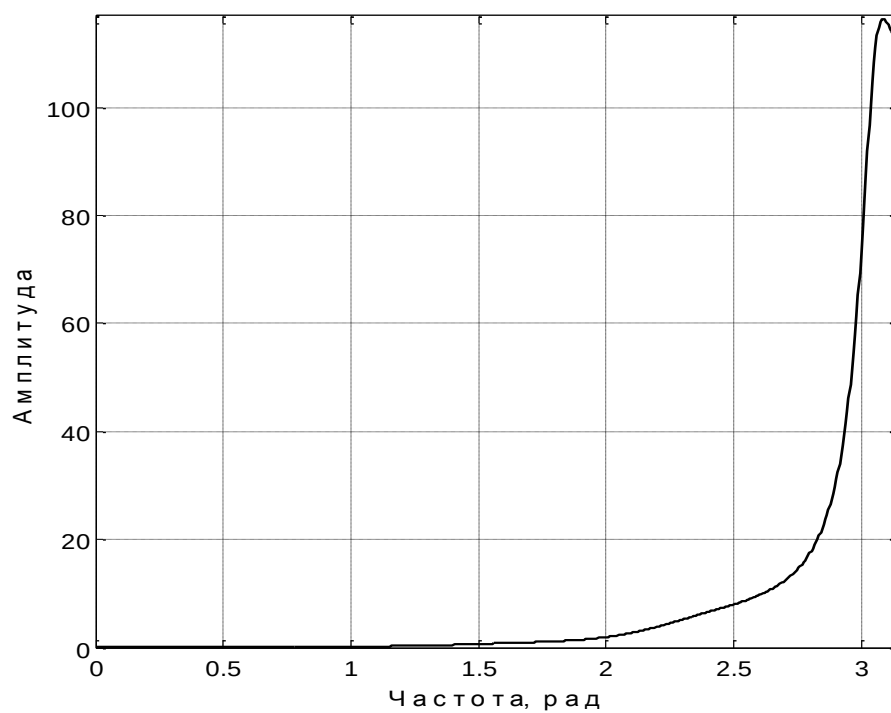


Рисунок 2.19 – АЧХ фильтра синтеза при условии, что нули функций $P'(\hat{\omega})$ и $Q'(\hat{\omega})$ принимают максимальные значения

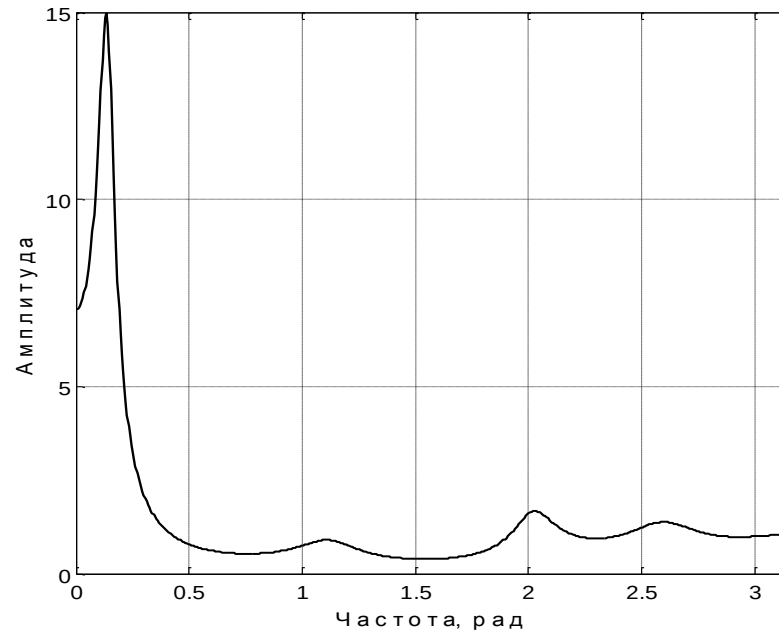


Рисунок 2.20 – Пример АЧХ фильтра синтеза полученной на записи русской речи

На (рисунке 2.18) и (рисунке 2.19) изображены амплитудно-частотные характеристики восстановленных фильтров синтеза. В качестве примера на (рисунке 2.20) приведена амплитудно-частотная характеристика фильтра синтеза, полученная на одном из сегментов записи русской речи. Из (рисунка 2.18) и (рисунка 2.19) видно, что форма АЧХ, и диапазон её значений в исследуемых случаях, являются нехарактерными для речи (рисунки 2.16, 2.17). Исходя из вышеизложенного следует, что условия, при которых значения нулей функций $P'(\hat{\omega})$ и $Q'(\hat{\omega})$ одновременно принимают или минимальные или максимальные значения на практике при обработке РС не встречаются. Реальное АЧХ голосового тракта, определяемое положением своих ЛСЧ, представлено на (рис. 2.20).

Необходимо отметить, что в реальных условиях РС зачастую в той или иной степени зашумлен [23]. Анализ условий обработки зашумленного РС показывает, что определение значений линейных спектральных частот выполняется с изначальной ошибкой, что в общем случае приводит к нестабильности процедуры их оценки и коррекции. Наличие данных фактов позволяет судить о необходимости предобработки РС с целью шумоподавления, что указывает на необходимость более детального рассмотрения основ анализа и обработки РС при шумоподавлении, что будет показано в следующем подразделе.

2.6 Выводы по второму разделу

1. В существующих низкоскоростных кодеках реализуется только параметрическая адаптация системы обработки, что ограничивает возможности по компрессии речи.

2. Перспективным направлением развития адаптивных КРС является разработка систем с переменной структурой, то есть систем, в которых реализуются процедуры структурно-параметрической адаптации.

3. Систему адаптивного структурно-параметрического кодирования РС целесообразно разрабатывать на основе следующих принципов:

- для прототипа выбирается класс CELP-кодеков РС, наиболее широко используемый в телекоммуникациях, имеющий широкие возможности для развития;
- адаптивную структурно-параметрическую перестройку кодека необходимо производить по входному сигналу путем анализа обрабатываемых сегментов РС;
- параметрическую адаптивную обработку целесообразно реализовывать по принципу "анализа через синтез";
- в соответствии с результатами анализа РС предлагается изменять длительность сегментов обработки РС, вычисляя и корректируя значений ЛСП на них.

4. Разработанная математическая модель адаптивной системы с переменной структурой и параметрами описывает этапы обработки РС в системах на основе CELP. В ней для кодирования РС заложены принципы, которые позволяют уменьшить скорость передачи при сохранении качественных показателей синтезированной речи.

5. Выделение однородных участков РС на длительности нескольких классических сегментов анализа дает возможность уменьшить информационный ресурс кадра передачи, используемый для представления данных о передаточной функции голосового тракта.

3. МОДЕЛЬ СИСТЕМЫ ОБРАБОТКИ РЕЧЕВОГО СИГНАЛА, УЧИТЫВАЮЩАЯ ЗАВИСИМОСТИ ЭЛЕМЕНТОВ ДЕКОМПОЗИЦИИ ПРИ КОДИРОВАНИИ

3.1 Вводные замечания

Использование принципов структурно-параметрической адаптации системы обработки РС и углубленный детальный анализ параметров ЛП при их кластерном представлении позволили разработать новые технические решения по созданию устройств низкоскоростного кодирования РС в системах связи.

В данном разделе рассмотрены теоретические основы присутствия взаимосвязи элементов декомпозиции РС при его обработке на основе метода ЛП (подраздел 3.2), разработана аналитическая модель системы обработки РС, учитывающая зависимости элементов его декомпозиции (подраздел 3.3), проведено обоснование возможности применения метода обработки РС, основанного на учете зависимости элементов декомпозиции (подраздел 3.4).

3.2 Выявление зависимости элементов декомпозиции речевого сигнала

Теоретически основополагающим базисом метода ЛП является авторегрессионная модель, применяемая для решения различных задач цифрового спектрального анализа и предполагающая в "идеальном" случае бесконечный порядок формирующей системы при возбуждении ее сигналом в виде дискретного белого гауссовского шума (рисунок 3.1). Ее идентификация связана с решением системы алгебраических матричных уравнений Юла–Уокера [116, 188].

В классической постановке задачи параметрического цифрового спектрального анализа возбуждение формирующего фильтра осуществляется сигналом $u(n)$, представляющим собой реализации дискретного белого шума с математическим ожиданием, равным нулю, и единичной дисперсией:

$$\begin{cases} M\{u(n)\} = 0, \\ D\{u(n)\} = \sigma^2\{u(n)\} = 1. \end{cases}$$

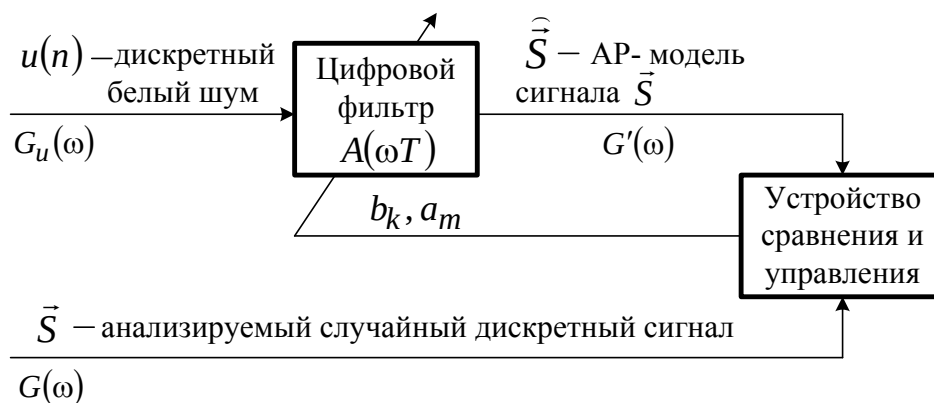


Рисунок 3.1 – Принцип параметрического спектрального оценивания на основе авторегрессионной модели сигнала

Точность идентификации математической модели исследуемого процесса напрямую связана с выбором величины ее порядка M . В качестве критерия настройки модели в предположении о гауссовском законе распределения исходного процесса используется среднеквадратическая ошибка:

$$e^2(n) = d_2(\vec{S}, \vec{S}') = \frac{1}{N} (\vec{S} - \vec{S}')^T (\vec{S} - \vec{S}') = \frac{1}{N} \sum_{i=1}^N (\vec{S} - \vec{S}')^2,$$

где $\vec{S}$ – вектор оригинального РС, $\vec{S}'$ – вектор синтезированного РС, N – количество отсчетов на сегменте анализа.

Повышение порядка передаточных функций фильтров анализа и синтеза приводит к "обелению" сигнала остатка предсказания, который является наилучшим сигналом возбуждения [5]. В классической постановке задачи параметрического цифрового спектрального анализа на основе авторегрессионной модели линейное разностное уравнение формирующего фильтра выглядит следующим образом:

$$y(nT) = - \sum_{m=1}^M a_m y(nT - mT) + u(nT),$$

где $y(nT)$ – выходной сигнал, T – интервал дискретизации, a_m – коэффициенты фильтра.

Его амплитудно-частотная характеристика определяется в виде

$$A(\omega T) = \frac{1}{\sqrt{(1 + \sum_{m=1}^M a_m \cos m\omega T)^2 + (\sum_{m=1}^M a_m \sin m\omega T)^2}}, \quad (3.1)$$

а спектральная плотность мощности

$$G(\omega) = \frac{\sigma^2 \{u(n)\} T}{(1 + \sum_{m=1}^M a_m \cos m\omega T)^2 + (\sum_{m=1}^M a_m \sin m\omega T)^2}, \quad (3.2)$$

где ω – круговая частота дискретного преобразования Фурье.

Повышение порядка модели в выражениях (3.1) и (3.2) приводит к получению более точных оценок относительно анализируемого сигнала $\vec{S}$. С точки зрения теории $e^2(n) \rightarrow 0$ при $M \rightarrow \infty$.

На (рисунке 3.2) представлена спектральная плотность мощности РС и сигнала остатка ЛП на сегменте квазистационарности. Данный рисунок получен для вокализованного сегмента РС на основе анализа остатка ЛП при $M=10$ фразы "Цифровая обработка сигналов" при $f_d=8$ кГц и 256 уровней квантования.

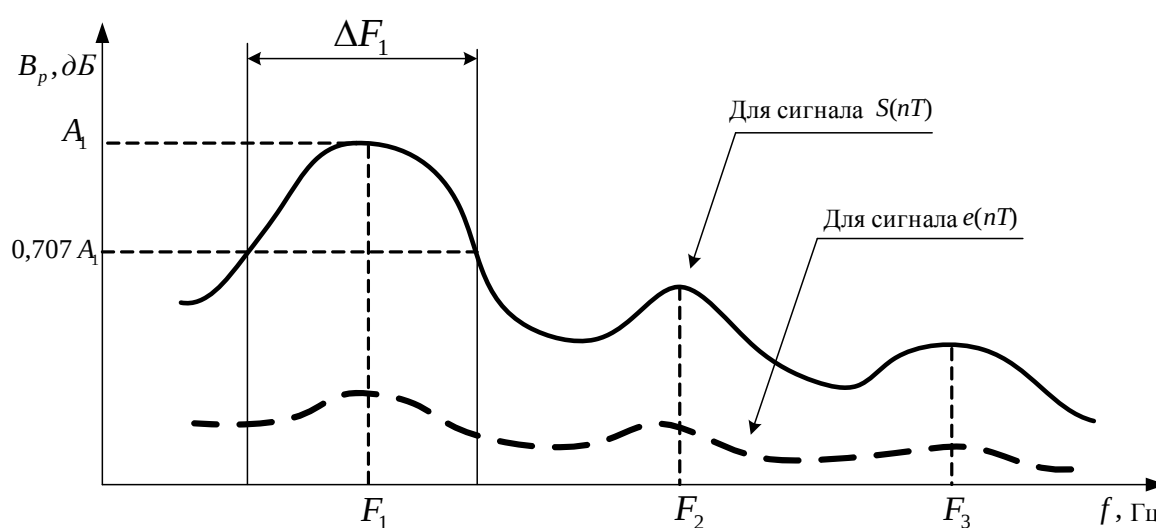


Рис.

Рисунок 3.2 – Оценка спектральной плотности мощности РС
и сигнала остатка линейного предсказания

На практике при реализации ЛП значение M всегда ограничено, что приводит к возникновению сигнала $e(n)$, являющегося сигналом возбуждения фильтра синтеза модели ЛП [5, 33, 42]. Таким образом, сигнал $e(n)$ не является реализациями белого шума с математическим ожиданием, равным нулю, и единичной дисперсией. Анализ (рисунка 3.2) показывает зависимости между поведением СПМ исходного сигнала и соответствующего сигнала ошибки. Сигнал $e(n)$ становится детерминированным относительно множества $\{a_m\}$ и связан с ним соответствующими зависимостями [43]. Таким образом, основываясь на использовании результатов анализа формантной структуры РС на сегменте анализа, существует возможность, опираясь на эти данные, формировать сигнал возбуждения синтезирующего фильтра [131].

3.3 Модель системы обработки речевого сигнала, учитывающая зависимость элементов его декомпозиции

Рассмотрим аналитическую модель системы обработки РС с учетом зависимости элементов его декомпозиции при кодировании (рисунок 3.3), где:

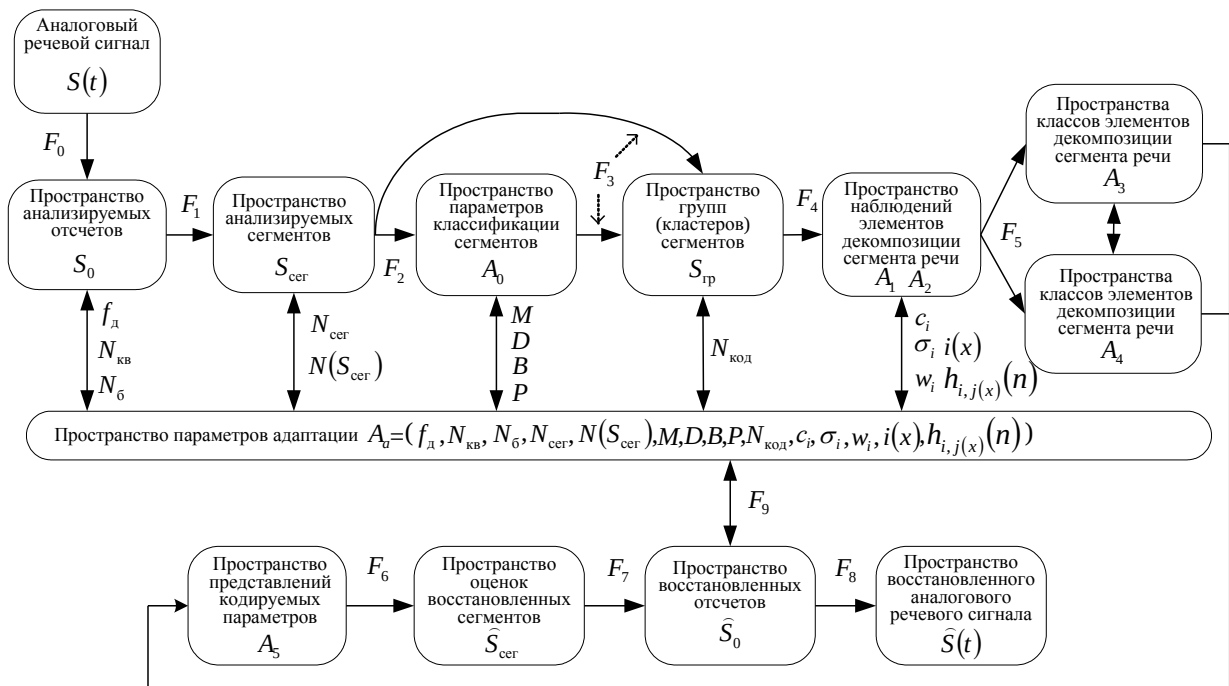


Рисунок 3.3 – Аналитическая модель системы обработки речевого сигнала, учитывающая зависимости элементов декомпозиции

1) $S(t)$ – пространство аналогового РС с ограниченным динамическим диапазоном и полосой частот. $f_n = 0,3 \text{ кГц}$, $f_b = 3,4 \text{ кГц}$.

2) F_0 – преобразование пространства непрерывного сигнала $S(t)$ в пространство анализируемых отсчетов S_0 . $F_0: \{S(t)\} \rightarrow \{S_0\}$.

F_0 описывает процедуру аналого-цифрового преобразования, в которой параметры f_d – частота дискретизации, $N_{\text{кв}}$ – число уровней квантования, N_6 – число бит кодирования могут быть постоянными величинами согласно рекомендаций МСЭ или адаптивными.

3) S_0 – пространство отсчетов: $S_0 = [S_1, S_2, \dots, S_N]$.

4) F_1 – преобразование пространства анализируемых отсчетов в пространство сегментов анализа $S_{\text{сег}}$. $F_1: \{S_0\} \rightarrow \{S_{\text{сег}}\}$.

5) $S_{\text{сег}}$ – пространство анализируемых сегментов. $S_{\text{сег}_n} = [S_1, S_2, \dots, S_n]$, где $S_{\text{сег}_n} \in \{S_{\text{сег}}\}$. $S_{\text{сег}_n}$ – анализируемый сегмент. Длина сегмента (количество отсчетов) $N_{\text{сег}}$ и число сегментов $N(S_{\text{сег}})$ могут адаптивно изменяться.

6) F_2 – извлечение из пространства анализируемых сегментов $S_{\text{сег}}$ пространства параметров классификации A_0 . Данное преобразование выполняет роль первого параметрического анализатора сегментов. $F_2: \{S_{\text{сег}}\} \rightarrow \{A_0\}$.

7) A_0 – пространство параметров классификации сегментов РС.

8) F_3 – отображение пространства анализируемых сегментов РС в пространство групп (кластеров) сегментов $S_{\text{гр}}$ с использованием пространства параметров классификации A_0 . $F_3: \{S_{\text{сег}}, A_0\} \rightarrow \{S_{\text{гр}}\}$.

9) $S_{\text{гр}}$ – пространство групп (кластеров) сегментов РС, $N_{\text{код}} < \infty$, $N_{\text{код}}$ – количество кластеров сегментов РС, $n_{\text{код}}$ – номер кластера, где $n_{\text{код}} \in (1; N_{\text{код}})$.

$S_{\text{гр-тон1}} - S_{\text{гр-тонN}}$ – подпространства тоновых сегментов РС;

$S_{\text{гр-пш1}} - S_{\text{гр-пшN}}$ – подпространства переходных и шумовых сегментов РС;

10) F_4 – отображение пространства групп (кластеров) сегментов $S_{\text{гр}}$ в пространство наблюдений элементов декомпозиции сегмента речи, коэффициентов формирующей модели и сигнала возбуждения. $F_4: \{S_{\text{гр}}\} \rightarrow \{A_1, A_2\}$.

11) A_1 – пространство наблюдений параметров идентификации фильтра анализа на передающей и синтеза на приемной стороне системы обработки.

$A_{\text{1гр-тон}} = [A_{\text{1гр-тон1}}, \dots, A_{\text{1гр-тонN}}]$, $A_{\text{1гр-пш}} = [A_{\text{1гр-пш1}}, \dots, A_{\text{1гр-пшN}}]$.

A_2 – пространство наблюдений сигнала возбуждения фильтра-синтезатора $A_{\text{2гр-тон}} = [A_{\text{2гр-тон1}}, \dots, A_{\text{2гр-тонN}}]$, $A_{\text{2гр-пш}} = [A_{\text{2гр-пш1}}, \dots, A_{\text{2гр-пшN}}]$

12) F_5 – отображение признаков пространств A_1 и A_2 в пространства наблюдений классов элементов декомпозиции сегмента речи. $F_5: \{A_1, A_2\} \rightarrow \{A_3, A_4\}$

13) $\{A_3, A_4\}$ – пространства классов элементов декомпозиции сегмента речи.

14) A_5 – пространство представлений кодируемых параметров.

- 15) F_6 – отображение пространства представлений кодируемых параметров A_5 в пространство оценок восстановленных сегментов РС $\hat{S}_{\text{сег}}$. $F_6: \{A_5\} \rightarrow \{\hat{S}_{\text{сег}}\}$
- 16) $\hat{S}_{\text{сег}}$ – пространство оценок восстановленных сегментов РС. $\hat{S}_{\text{сег}} C \in \{1; N_{\text{кадр}}\}$, где $N_{\text{кадр}} < \infty$. $N_{\text{кадр}}$ – общее число сегментов.
- 15) F_7 – отображение пространства восстановленных сегментов РС $\hat{S}_{\text{сег}}$ в пространство восстановленных отсчетов $\hat{S}_0$. $F_7: \{A_6\} \rightarrow \{\hat{S}_0\}$. Математически F_7 описывает формирование единого цифрового потока синтезированного РС.
- 18) $\hat{S}_0$ – пространство восстановленных отсчетов РС.
 $\hat{S}_0 = [\hat{S}_1, \hat{S}_2, \dots, \hat{S}_N]$, $n \in \{1; R\}$, где $R < \infty$.
- 19) F_8 – отображение пространства восстановленных отсчетов $\hat{S}_0$ в пространство восстановленного аналогового РС $\hat{S}(t)$: $F_8: \{\hat{S}_0\} \rightarrow \{\hat{S}(t)\}$.
- 20) $\hat{S}(t)$ – пространство восстановленного аналогового РС с ограниченными динамическим диапазоном и полосой частот.
- 21) F_9 – отображение пространства анализируемых отсчетов S_0 и пространства параметров адаптации A_a в пространство восстановленных отсчетов РС $\hat{S}_0$.

$$F_9: \{S_0, A_a\} \rightarrow \{\hat{S}_0\}.$$

Пространство параметров адаптации A_a имеет вид:

$$A_a = \{f_d, N_{\text{кв}}, N_{\text{б}}, N_{\text{сег}}, N(S_{\text{сег}}), M, D, P, N_{\text{код}}, C_i, \sigma_i, w_i, i(x), h_{i,j(x)}(n), w_i\}$$

Использование зависимостей элементов декомпозиции РС при ЛП позволяет по-новому рассмотреть процесс обработки РС [139, 143], поскольку большинство работ в этой области использовали независимое представление и кодирование пространств представлений данных параметров [29]. В (табл. 3.1) представлено распределение информационных ресурсов в стандартизованных липредерах. Анализ (табл. 3.1) позволяет сделать вывод о том, что основная доля информационных ресурсов приходится на представление сигнала возбуждения.

Систему низкоскоростного кодирования РС можно рассматривать как некоторый архиватор РС, работающий в реальном масштабе времени "на лету". Низкоскоростные системы обработки РС в этом контексте являются архиваторами с высоким уровнем уплотнения информации. Нахождение сигнала возбуждения является для них примерно в 4 раза более ресурсоемкой задачей в сравнении с более

простой задачей идентификации и передачи параметров формирующей модели. Чем выше коэффициент сжатия речевой информации, тем больше внимания уделяется идентификации сигнала возбуждения и формированию его архива в виде векторов кодовой книги (КК), основная доля выделенного информационного ресурса кадра передачи тратится именно на его представление (таблица 3.1).

Таким образом, учет выявленных зависимостей и их использование при совершенствовании устройств обработки РС приводит к снижению средней скорости передачи (кодирования) при его обработке [10,14,42,47,48,51,52,53].

Зависимости между параметрами, описывающими передаточную функцию голосового тракта на сегменте анализа, и соответствующим сигналом возбуждения объясняется особенностями постановки и решения обратной задачи ЦСА при фиксированном порядке формирующего фильтра [117].

Таблица 3.1 – Распределение информационных ресурсов
в стандартизованных липредерах

Стандарт	GSM	G.729	Усредненное распределение
Тип липредера	RPE-LTP -LPC	CS- ACELP	
$V_K, кбит / с$	13	8	
Кодируемые параметры	Относительные информационные затраты (скорость кодирования параметра/ V_K)		
Коэффициенты модели	0,14	0,23	0,24
Сигнал возбуждения	0,86	0,77	0,76

При формировании ограниченных множеств параметров голосового тракта и сигналов возбуждения в виде КК данные зависимости вырождаются в соответствующие взаимосвязи классов подпространств соответствий между собой, что показано на (рисунок 3.4).

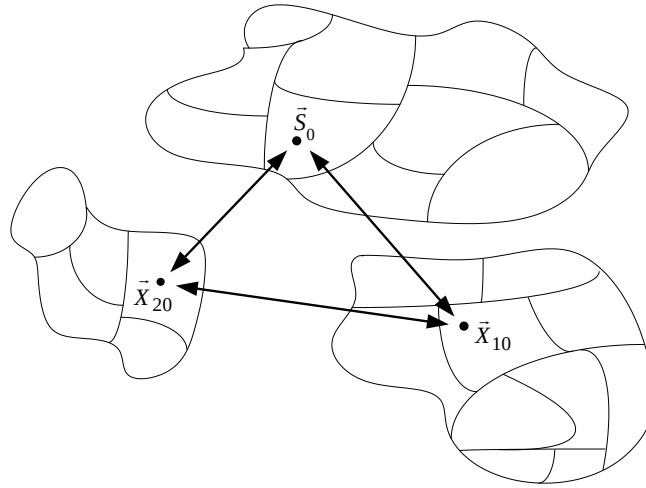


Рисунок 3.4 – Классы подпространств представлений
соответствий элементов декомпозиции РС

С учетом рассмотренной математической модели (рисунок 3.3) реализация данных зависимостей требует динамического перераспределения мощностей подпространств $\{Y_i\}$ пространства представлений кодируемых параметров с возможным изменением отображений F_{2i} [52]. Такая задача относится к классу задач эффективного кодирования, а в условиях идеального КС – к классу задач сжатия РС. На (рисунке 3.5) и (рисунке 3.6) представлены структурная схема кодера и математическая модель кодирования РС.

К адекватным способам определения правильности функционирования синтезируемой системы кодирования относится имитационное моделирование. Для упрощения процедуры моделирования и многовариантного анализа систем применяют проблемно-ориентированные прикладные программы технических расчетов, одной из которых является MATLAB [26, 30, 32, 88, 93, 241]. Особенностью схемы (рисунок 3.5) является введение процедуры классификации сегментов РС на непесекающееся число классов. Учет зависимости элементов декомпозиции РС требует совершенствования устройств параметрического линейного предсказания с ВК, используемых в существующих стандартах речепреобразования. В настоящее время они представляют собой системы обработки РС на основе ЛП с ВК при ко-

нечном числе состояний, определяемых размерами КК. Использование данных особенностей позволяет оптимизировать системы ЛП с ВК, снижая существующую избыточность фиксированных КК.

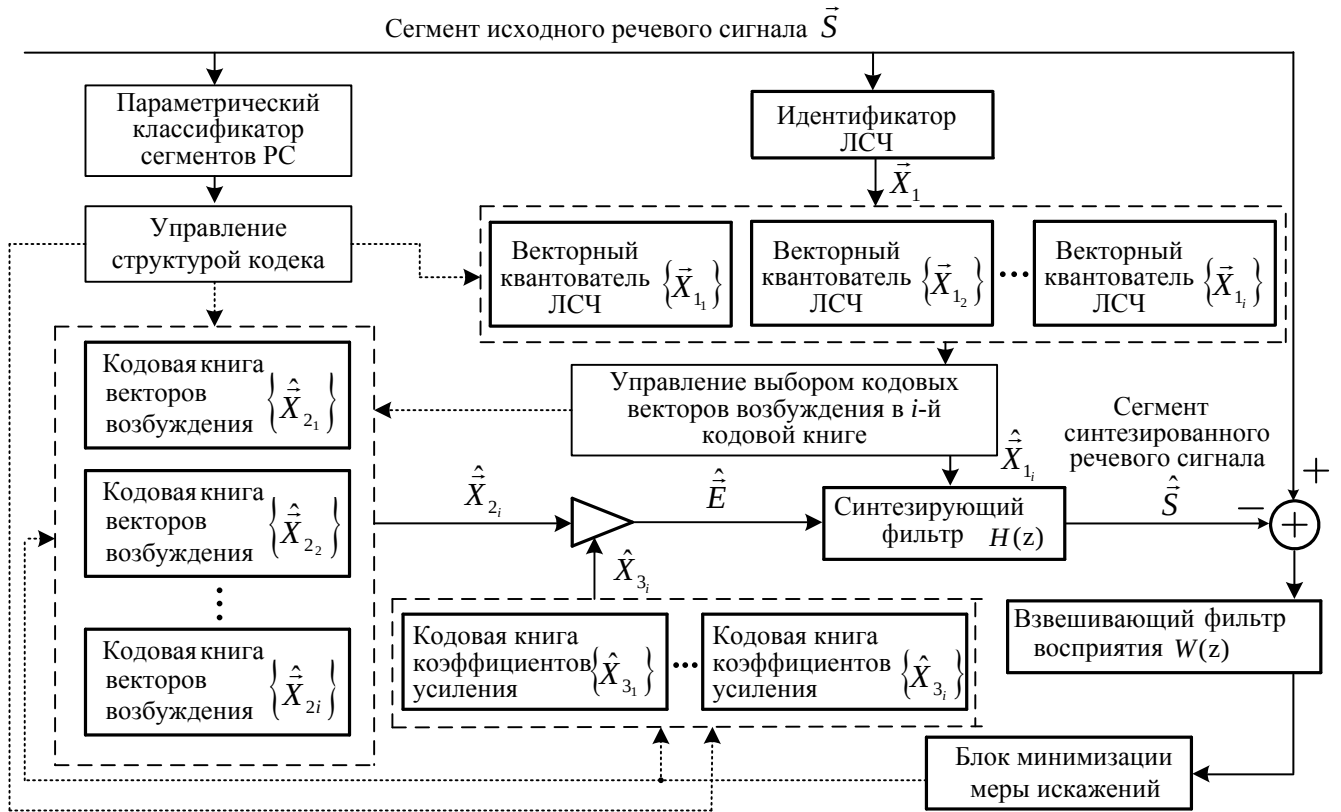


Рисунок 3.5 – Структурная схема кодера речевого сигнала с учетом зависимостей элементов декомпозиции

Первым отличием предложенной математической модели от представленной в разделе 2, является введение процедуры параметрической классификации сегментов РС, что отражается в формировании пространства состояний кодера $\{\bar{C}_x\}$ из пространства анализируемых сегментов $\{\bar{S}\}$ посредством отображения F_a .

$$F_a: \{\bar{S}\} \rightarrow \{\bar{C}_x\}. \quad (3.3)$$

Смысловым содержанием F_a выступает процедура параметрического анализа входных сегментов РС, рассмотренная в [17, 44, 186, 188], а эффективность реализации предлагаемого подхода зависит от вида F_a и $\{\vec{S}\}$.

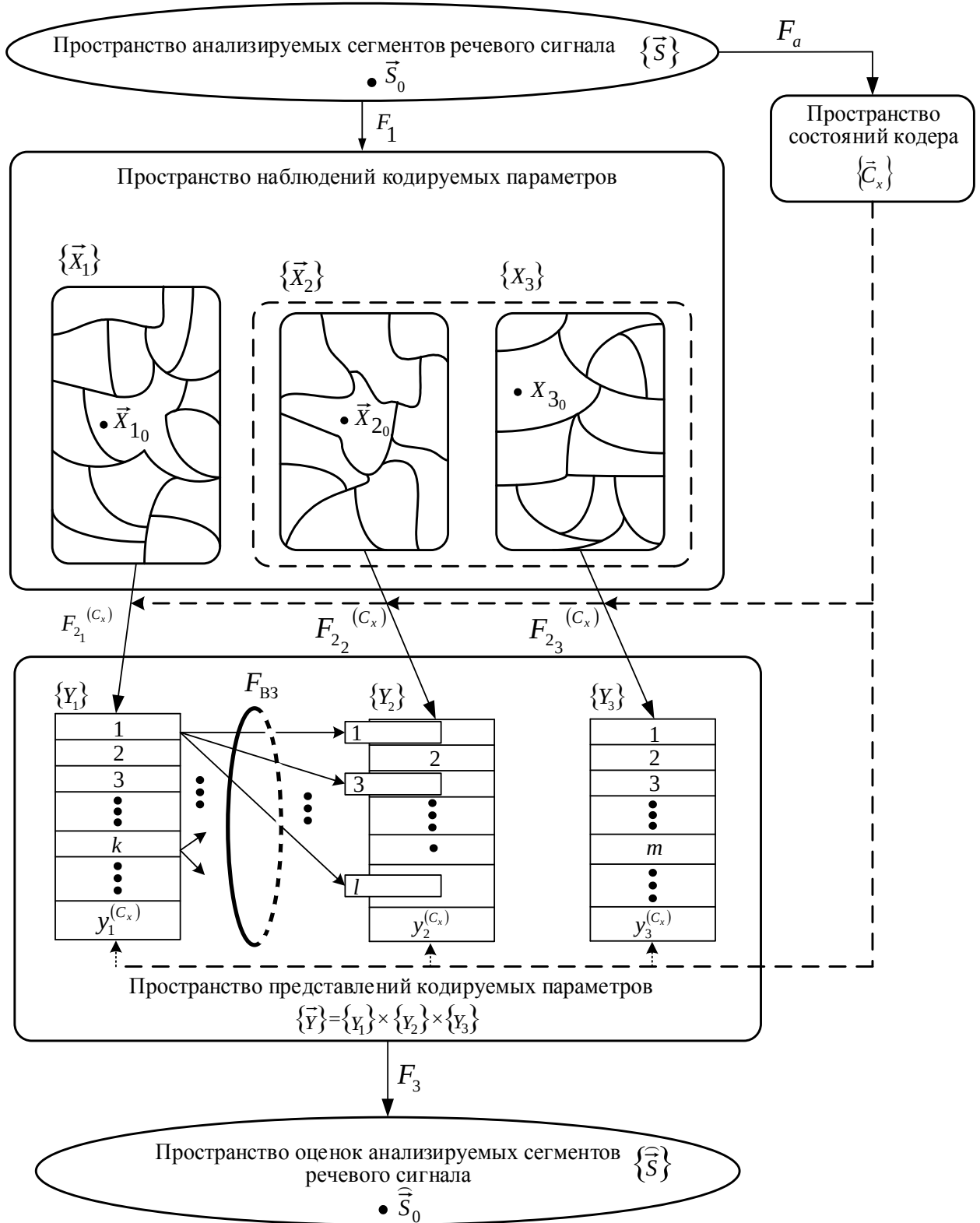


Рисунок 3.6 – Математическая модель процедур преобразования РС в системе гибридного кодирования с учетом зависимостей элементов декомпозиции

Особенностью предлагаемой системы кодирования РС является необходимость передачи по КС служебной информации о состоянии кодера на текущем сегменте анализа РС. Используемые в современных гибридных кодерах отображения F_{2_i} оптимизируют кодеки в условиях постоянства мощностей подпространств $\{Y_i\}$; а в разработанной системе необходим совместный поиск значений y_i и отображений F_{2_i} . В связи с этим множество состояний $\{\bar{C}_x\}$ кодера представляет собой декартово произведение:

$$\{\bar{C}_x\} = \{No_{F_{2_1}}\} \times \{No_{F_{2_2}}\} \times \dots \times \{No_{F_{2_i}}\} \times \{y_1\} \times \{y_2\} \times \dots \times \{y_i\}, \quad (3.4)$$

где $No_{F_{2_i}}$ – номер возможного отображения F_{2_i} , $y_i = \overline{1, \gamma}$, $i = \overline{1, 3}$.

Исключение i – х подпространств из структуры пространства представлений кодируемых параметров обеспечивается при $y_i = 1$. При однозначном соответствии значений y_i и отображений F_{2_i} множество $\{\bar{C}_x\}$ определяется как $\{\bar{C}_x\} = \{y_1\} \times \{y_2\} \times \dots \times \{y_i\}$.

При заданной величине γ и выделении целого числа бит для кодирования каждого подмножества $\{Y_i\}$ область изменения y_i представляет собой конечное множество целых чисел степени «2», т. е. $y_i = \{2^0, 2^1, \dots, \gamma\}$.

Количество возможных векторов состояний кодера при $\gamma = \prod_{i=1}^3 y_i$ и однозначном соответствии отображений F_{2_i} значениям y_i с выделением целого числа бит на кодирование каждого подмножества $\{Y_i\}$ определяется количеством сочетаний с повторениями из 3 по $\log_2 \gamma$

$$\alpha = C_3^{\log_2 \gamma} = \frac{(2 + \log_2 \gamma)!}{2! \log_2 \gamma!}.$$

Реализация такого подхода требует, чтобы множество $\{\vec{C}_x\}$ было конечным, а его мощность α определялась на основании компромисса между получаемым выигрышем при оптимизации и требуемыми затратами на информационный ресурс при передаче данных о текущем состоянии кодера ($\log_2 \alpha$). При таких условиях возможно использование алгоритма, допускающего несколько режимов работы кодера, соответствующих конечному числу непересекающихся классов сегментов РС.

Второе отличие – учет зависимостей параметров голосового тракта ЛСЧ и сигнала возбуждения – обуславливает введение оператора $F_{ВЗ}$, реализующего процедуру выбора или установления соответствия между вектором ЛСЧ и конечным набором векторов возбуждения из i -й КК сигналов возбуждения

$$F_{ВЗ} : \{Y_1\} \rightarrow \{Y_{2_i}\}. \quad (3.5)$$

При однозначном соответствии вектора ЛСЧ и группы векторов возбуждения (представленной как z_i из i -й КК) количество возможных векторов состояния кодера при выборе соответствующих КК для параметров голосового тракта и сигнала возбуждения представляется числом сочетаний с повторениями из 1 по z_i .

$$\beta = C_1^{z_i} = \frac{(1 + z_i)!}{1! z_i!}$$

Следствием второго отличия выступает особенность представления вектора $\vec{Y}$. Координаты $\vec{X}_{1_0}$ и $\vec{X}_{2_0}$ данного случайного вектора считаются зависимыми, следовательно, их совместная плотность распределения имеет вид

$$W(\vec{X}_{1_0}, \vec{X}_{2_0}, X_{3_0}) = W_1(\vec{X}_{1_0}) \cdot W_2(\vec{X}_{2_0} | \vec{X}_{1_0}) \cdot W_3(X_{3_0}). \quad (3.6)$$

С учетом случайного характера вектора $(\vec{X}_{1_0}, \vec{X}_{2_0}, X_{3_0})$ есть возможность усреднить величину искажения РС по всем анализируемым речевым векторам, отображаемым в вектор $\vec{Y}$:

$$D(\vec{S}_0, \hat{\vec{S}}_0) = M[D(\vec{S}_0, \hat{\vec{S}}_0)] = \int_{\{\vec{X}_{10}\}} \int_{\{\vec{X}_{20}\}} \int_{\{\vec{X}_{30}\}} D(\vec{S}_0, \hat{\vec{S}}_0) W(\vec{X}_{10}) \cdot W(\vec{X}_{20} | \vec{X}_{10}) \cdot W(X_{30}) \times \\ \times d\vec{X}_{10} (d\vec{X}_{20} | \vec{X}_{10}) dX_{30}. \quad (3.7)$$

Полученная величина $D(\vec{S}_0, \hat{\vec{S}}_0)$ характеризует среднее значение избранной меры искажения РС при обработке последнего алгоритмом кодирования на основе метода ЛП.

При классификации сегментов РС на $\{\vec{C}_x\}$ непересекающихся классов с характеристиками $c_x = 1, 2, \dots, C_x$ – номер класса, $P^{(c_x)}$ – вероятность принадлежности текущего сегмента РС к классу c_x при $\sum_{c_x=1}^{C_x} P^{(c_x)} = 1$; $W^{(c_x)}(\vec{X})$ – плотность распределения вероятности вектора $\{\vec{X}\}$ в классе c_x ; $D^{(c_x)}(\vec{S}_0, \hat{\vec{S}}_0)$ – средняя величина выбранной меры искажения для сегментов РС класса c_x , зависимость математического ожидания имеет вид

$$\overline{D}(\vec{S}_0, \hat{\vec{S}}_0) = M[D^{(c_x)}(\vec{S}_0, \hat{\vec{S}}_0)] = \sum_{c_x=1}^{C_x} P^{(c_x)} D^{(c_x)}(\vec{S}_0, \hat{\vec{S}}_0). \quad (3.8)$$

В случае конечности множества состояний кодера $\{\vec{y}_i^{(c_x)}\}$ мощности $\alpha^{(c_x)}$ можно считать состояния кодера, принадлежащие этому классу сегментов, равновероятными: $P(\vec{y}_\theta^{(c_x)}) = \frac{1}{\alpha^{(c_x)}}$, где $\theta = 1, 2, \dots, \alpha^{(c_x)}$ – номер текущего состояния кодера для класса C_x .

При данном способе формирования множества состояний кодера его мощность будет определяться как

$$\alpha = \sum_{c_x=1}^{C_x} \alpha^{(c_x)}. \quad (3.9)$$

Для C_x -го класса сегментов средняя величина D определяется как

$$D^{(c_x)}(\vec{S}_0, \hat{\vec{S}}_0) = \frac{1}{\alpha^{(c_x)}} \sum_{\theta=1}^{\alpha^{(c_x)}} D(\vec{S}_0, \hat{\vec{S}}_0). \quad (3.10)$$

Среднее искажение кодируемых кадров РС при обработке будет определяться выражением

$$D(\vec{S}_0, \hat{\vec{S}}_0) = M \left[D^{(C_x)}(\vec{S}_0, \hat{\vec{S}}_0) \right] = \sum_{c_x=1}^{C_x} P^{(C_x)} D^{(C_x)}(\vec{S}_0, \hat{\vec{S}}_0) = \sum_{c_x=1}^{C_x} P^{(C_x)} \frac{1}{\alpha^{(C_x)}} \times$$

$$\times \sum_{\theta=1}^{\alpha^{(C_x)}} \frac{1}{\gamma} \sum_{Y_1=1}^{y_1} \sum_{Y_2=1}^{y_2} \sum_{Y_3=1}^{y_3} D\{\vec{X}_0, F_3[F_{2_1}(\vec{X}_{10}), F_{2_2}(\vec{X}_{20}), F_{2_3}(\vec{X}_{30})]\} \times P_W(\vec{X}_{10}, \vec{X}_{20} | \vec{X}_{10}, \vec{X}_{30}) \quad (3.11)$$

Практически значимой особенностью разрабатываемой системы является возможность построения гибридных кодеров РС учитывающих зависимость элементов декомпозиции РС в процессе его обработки на основе метода ЛП, использующих ВК данных параметров [34].

Параметры кодека, оказывающие доминирующее влияние на качество кодирования РС, играют роль переменных при решении задачи разработки системы кодирования и определяют значение целевой функции вида

$$D(\vec{S}_0, \hat{\vec{S}}_0) \rightarrow_{\vec{\varepsilon}} \min, \quad \vec{\varepsilon} = \{C_x, z^{(C_x)}_i, \alpha^{(C_x)}, M, y_i^{(C_x)}\},$$

где C_x – конечное число непересекающихся классов сегментов РС;

$z^{(C_x)}_i$ – число векторов возбуждения, зависящих с вектором ЛСЧ;

$\alpha^{(C_x)}$ – мощность множества состояний кодера для C_x -го класса сегментов РС;

M – количество коэффициентов ЛП;

$y_i^{(C_x)}$ – объем i -ой КК для C_x -го класса сегментов РС.

В случае кодирования РС без учета служебной информации аналитическое выражение, определяющее скорость кодирования РС, может быть представлено следующим образом:

$$V_k = \frac{\left(\log_2 \prod_{i=1}^3 y_i + \log_2 C_x + g \right) \cdot f_d}{N} \text{ кбит/с}, \quad (3.12)$$

где g – количество бит синхронизации.

3.4 Метод обработки речевого сигнала с учетом зависимости элементов его декомпозиции

Представленная математическая модель системы гибридного кодирования РС с учетом зависимостей элементов его декомпозиции позволяет разработать алгоритмы кодирования РС. Использование зависимости элементов декомпозиции РС $\{\bar{a}\}$ и $\{\bar{e}\}$ при создании аналитической модели потребует формирования кластеризованных пространств $\{\bar{a}\}$, $\{\bar{e}\}$ и нахождение взаимосвязи между ними.

Введем утверждение: при ограниченном порядке ЛП РС векторные множества элементов декомпозиции РС (вектора параметров формирующей модели $\{A\}$ и сигнала возбуждения $\{E\}$) становятся зависимыми и определяются многозначными необратными отображениями между соответствующими векторными множествами.

Для его доказательства воспользуемся методом от противного в рамках положений метода порождающего процесса в предположении, что сигнал остатка (возбуждения) ЛП представляет собой дискретный белый гауссовский шум (ДБГШ) в соответствии с моделью авторегрессионного цифрового спектрального анализа, то есть его реализации не коррелированы между собой и не связаны с векторами параметров формирующей модели [117]. Тогда РС может быть представлен выражением (3.13), при этом его оценка на интервале nT есть выражение (3.14)

$$S(nT) = S'(nT, (\bar{a}, \bar{e})) + n_0(nT), \quad (3.13)$$

$$\hat{S}'(nT) = M \{S'(nT, (\bar{a}, \bar{e})) | S_0(nT)\} = \sum_{nT} S'(nT, (\bar{a}, \bar{e})) p(nT, (\bar{a}, \bar{e})), \quad (3.14)$$

где $S'(nT, (\bar{a}, \bar{e}))$ – предсказанный РС; $\hat{S}'(nT)$ – оценка предсказанного РС; $n_0(nT)$ – собственный аппаратный шум.

$$p(nT, (\bar{a}, \bar{e})) = p(nT, (\bar{a}, \bar{e}) | S_0(nT)). \quad (3.15)$$

где $p(nT, (\bar{a}, \bar{e}))$ – апостериорная плотность вероятности параметров его декомпозиции на интервале nT .

Порождающий процесс для сигнала, представленного выражением (3.13), представим через (3.16), где $\xi(nT)$ – ДБГШ, следовательно, должны выполняться

условия (3.17).

$$\xi(nT) = S(nT) - \hat{S}'(nT), \quad (3.16)$$

$$M\{\xi(nT)\} = 0; \quad M\{\xi(n_1T)\xi(n_2T)\} = N_0\delta(n_1T - n_2T)/2, \quad (3.17)$$

где N_0 – односторонняя спектральная плотность мощности ДБГШ.

$$\begin{aligned} M\{\xi(nT)\} &= M\{S(nT)\} - M\{\hat{S}'(nT)\} = M\{S(nT, (\vec{a}, \vec{e})) + n_0(nT)\} - M\{\hat{S}'(nT)\} = \\ &= M\left\{M\{S'(nT, (\vec{a}, \vec{e}))|S_0(nT)\}_{(\vec{a}, \vec{e})}\right\}_{S_0(nT)} - M\{\hat{S}'(nT)\} = M\{\hat{S}'(nT)\} - M\{\hat{S}'(nT)\} = 0. \end{aligned} \quad (3.18)$$

Для доказательства второго положения выражения (3.17), введем иное задание порождающего процесса через выражения (3.19), (3.20):

$$\xi(nT) = \varepsilon(nT) + n_0(nT), \quad \varepsilon(nT) = S'(nT, (\vec{a}, \vec{e})) - \hat{S}'(nT), \quad (3.19)$$

$$\begin{aligned} M\{\xi(nT)\xi(nT + \Delta T)\} &= M\{\xi(nT)[\varepsilon(nT + \Delta T) + n_0(nT + \Delta T)]\} = \\ &= M\{\xi(nT)\varepsilon(nT + \Delta T)\} + M\{\xi(nT)n_0(nT + \Delta T)\} + M\{n_0(nT)n_0(nT + \Delta T)\}. \end{aligned} \quad (3.20)$$

В выражении (3.20) первые два слагаемых можно считать равными нулю, что подтверждается (3.21), (3.22), таким образом (3.20) трансформируется в (3.23), что свидетельствует о том, что $\xi(nT)$ является дискретным белым шумом.

$$\begin{aligned} M\{\xi(nT)\varepsilon(nT + \Delta T)\} &= M\left\{M\{\xi(nT)\varepsilon(nT + \Delta T)/S_0(nT + \Delta T)\}_{(\vec{a}, \vec{e})}\right\}_{S_0(nT + \Delta T)} = \\ &= M\left\{M\{\varepsilon(nT + \Delta T)/S_0(nT + \Delta T)\}_{(\vec{a}, \vec{e})}\xi(nT)\right\}_{S_0(nT)} = 0 \text{ при } \Delta T > 0 \end{aligned} \quad (3.21)$$

$$\text{При } \hat{S}'(nT + \Delta T) \text{ справедливо } M\{\varepsilon(nT + \Delta T)/S_0(nT + \Delta T)\} = 0$$

$$\begin{aligned} M\{\varepsilon(nT)n_0(nT + \Delta T)\} &= M\{[S'(nT, (\vec{a}, \vec{e})) - \hat{S}'(nT)]n_0(nT + \Delta T)\} = \\ &= M\{S'(nT, (\vec{a}, \vec{e}))\}M\{n_0(nT + \Delta T)\} - M\{\hat{S}'(nT)\}M\{n_0(nT + \Delta T)\} = 0. \end{aligned} \quad (3.22)$$

Порождающий процесс $\xi(nT)$ представлен (3.19) ($\varepsilon(nT)$ – ошибка оценки $S'(nT, (\vec{a}, \vec{e}))$ с распределением $p_\varepsilon(\varepsilon, nT)$) при выполнении условия (3.23), следствием из которого является (3.24), так как $\varepsilon(nT)$ и $n(nT)$ некоррелированы при $\Delta T > 0$, и плотность вероятности их суммы равна свертке "широкой" нормальной плотности вероятности и "узкой" плотности вероятности оценки процесса.

$$D\{n_0(nT)\} \gg D\{\varepsilon(nT)\}, \quad (3.23)$$

$$M\{\xi(nT)\xi(nT + \Delta T)\} = M\{\xi(nT)\xi(nT + \Delta T)\} = M\{n_0(t)n_0(t + \Delta T)\} = N_0\delta(\Delta T)/2 \quad (3.24)$$

Таким образом, формируется оценка (3.25)

$$p_\xi(n_\xi, nT) = \sum_{\varepsilon} p_0(\xi - \varepsilon; nT) p_\varepsilon(\varepsilon, nT) \approx p_0(\xi, nT). \quad (3.25)$$

Из (3.25) следует, что $\xi(nT)$ должен быть случайным процессом вида ДБГШ, однако анализ сигнала $e(n)$ (рис. 3.2) позволяет утверждать, что он не является процессом вида ДБГШ.

Анализ спектральной плотности мощности РС и ошибки ЛП (рис. 3.2) показывает взаимосвязь данных сигналов, что подтверждает зависимость сигнала остатка предсказания и коэффициентов синтезирующей РС модели.

Таким образом, решается задача совместного оценивания параметров формирующей модели и порождающего сигнала. Зачастую оценки порождающего процесса, которые могут быть синтезированы путем применения механизма обратной фильтрации в процедуре цифрового спектрального анализа, не соответствуют предположениям о реальном возбуждении. При рассмотрении порождающего сигнала в виде некоррелированной последовательности отсчетов согласно [116] модель обработки РС представляется линейной. Однако, если используется информация о параметрах элементов декомпозиции РС с использованием ВК, то даже на длительности сегмента анализа формирующая модель является нелинейной [125]. Такое представление также дает возможность использования методов непараметрической статистики для оценок распределения параметров декомпозиции РС, которые используются при кодировании. При этом нет необходимости априорного знания аналитических распределений плотностей вероятности элементов декомпозиции РС.

На основании вышеизложенного сформулируем следующее:

Утверждение 2. Учет зависимости элементов множеств декомпозиции речевого сигнала, получаемых на основе модели линейного преобразования, между пространствами кластеризации при его моделировании, приводит к уменьшению мощности используемого множества представления одного из векторных пространств при обработке.

На основании данного утверждения в четвертом разделе разработан алгоритм уменьшения избыточности представления РС в процессе его обработки и низкоскоростного кодирования для передачи по каналам цифровой связи. Основная трудность практической реализации системы кодирования с учетом зависимостей связана с отсутствием априорной информации о вероятной мощности пространства представлений различных вариантов КК. Устранение данной неопределенности частично реализуется при рациональном ограничении возможных вариантов при заданном качестве обработки, так как необходимо передавать информацию о состоянии и структуре системы в процессе ее функционирования. При совершенствовании систем обработки РС многие решения носят эмпирический характер, так как пока отсутствует четкий математический аппарат, который позволил бы связать воедино паритетный выбор между количеством возможных структур, качеством функционирования, алгоритмом работы и сложностью.

В общем случае при обработке РС и использовании его статистических зависимостей в идеальном виде требуется знание ожидаемой ошибки $e(n)$ от принятого решения U и значений $S(n)$ обрабатываемого РС. Для этого требуется статистическое описание наблюдений $S(n)$ а также параметров состояния синтезирующей системы, определяемых значениями $(\vec{a}, \vec{e})$. При этом полезный объем таких данных определяет значение ожидаемой ошибки при любом из ограниченного множества возможных решений. Решение задачи разработки системы в такой постановке осуществляется при известных распределениях вероятности $p(S|(\vec{a}, \vec{e}))$ и $p(\vec{a}, \vec{e})$.

На практике полное статистическое описание $S(n)$ и $(\vec{a}, \vec{e})$ получить невозможно, что связано с высокой размерностью решаемой задачи и сложностью статистического анализа речевых данных. При этом задачи обработки речевых данных

сопровождаются большей или меньшей априорной неопределенностью, которая ограничивает полноту статистического описания. Можно аппроксимировать распределения $S(n)$ и $(\vec{a}, \vec{e})$ с использованием нормального распределения [77], однако такой подход носит достаточно приближенный характер, ограничивающий область применения решений, и привносит ошибки в конечные результаты вычислений.

Таким образом, существуют затруднения в прямом использовании байесова подхода со всеми его преимуществами и возможностями [41, 98]. Возникающая априорная неопределенность относительно свойств источника делает особо важной разработку методов разработки таких систем информации и принятия решений при таких условиях.

Значения $(\vec{a}, \vec{e})$ носят векторный характер, они связаны друг с другом и со значениями $S(n)$ некоторыми функциональными ограничениями и зависимостями. Мощности пространств представлений параметров $\{\vec{a}\}$ и $\{\vec{e}\}$ в общем случае различны. При этом для получения приемлемых результатов обработки значения $(\vec{a}, \vec{e})$ неизменны на всем интервале времени для v , где v – порядковый номер сегмента анализа.

При наличии таких ограничений общее выражение для функции ошибки определяется в виде $e(\{U\}, \vec{a}, M, S)$ при наличии функции правдоподобия $p(S | (\vec{a}, \vec{e}))$. При таком подходе отражаются некоторые статистические характеристики $\{\vec{a}\}$ и $\{\vec{e}\}$, на основании которых вычисляется взаимосвязь между ними. Для задания подробного статистического описания в данном случае необходимы математическое ожидание параметров, дисперсия, а также моменты более высокого порядка. Априорные данные могут быть получены путем задания некоторого математического ожидания многокомпонентной функции $f(\vec{e})$. С теоретической и практической точки зрения большое значение имеет вычисление корреляционной матрицы $R = \|B\|$ вектора $\vec{e}$. Результаты таких вычислений могут использоваться для управления мощностью пространства представлений сигнала возбуждения.

При этом компоненты векторов $(\vec{a}, \vec{e})$ могут рассматриваться как векторный марковский случайный процесс, т. е.

$$p(\vec{a}, \vec{e}) = \prod_{i=2}^V p_1[(\vec{a}, \vec{e})_i | (\vec{a}, \vec{e})_{i-1}] p_0(\vec{a}, \vec{e}),$$

где переходная $p_1[(\vec{a}, \vec{e})_i | (\vec{a}, \vec{e})_{i-1}]$ и начальная $p_0(\vec{a}, \vec{e})$ плотности в общем случае могут быть полностью неизвестны.

Степень такой эквивалентности зависит от объема эмпирических данных, которые могут быть использованы по-разному: непосредственно для нахождения недостающих распределений вероятности; для оценки функциональной зависимости апостериорного риска от $S(n)$ и решения относительно $(\vec{a}, \vec{e})$ с целью уточнения структуры и параметров решающего правила, т. е. алгоритма обработки данных наблюдения $S(n)$.

Таким образом, вместо единственного распределения вероятности для параметров $(\vec{a}, \vec{e})$ с плотностью $p(\vec{a}, \vec{e})$ для решения задач обработки РС задается класс распределений P_0 , к которому относятся все возможные распределения $p(\vec{a}, \vec{e}) \in P_0$, что на практике выполнить достаточно сложно.

Наиболее универсальным способом статистического описания априорной неопределенности и учета имеющихся ограниченных сведений применительно к наблюдениям $S(n)$ и к ненаблюдаемым параметрам $(\vec{a}, \vec{e})$, является параметрический способ.

Двумерное векторное множество $(\vec{a}, \vec{e})$ дискретно и каждое его значение принимается с определенной вероятностью. Если априорное распределение вероятностей неизвестно, то в качестве неизвестных параметров рассматриваются сами эти вероятности, формальное задание которых с учетом естественных ограничений

$$p_i(\vec{a}, \vec{e}) \geq 0, \sum_{k=1}^P p_k = 1 \text{ определяет необходимое распределение вероятностей через}$$

$P-1$ неизвестный параметр.

Последовательность переходов значений $(\vec{a}, \vec{e})$ при условии распределения по нормальному закону определяет Марковский процесс:

$$p_1((\vec{a}, \vec{e})_v | (\vec{a}, \vec{e})_{v-1}) = p_1((\vec{a}, \vec{e})_v | (\vec{a}, \vec{e})_{v-1}, a, \sigma^2) = \frac{1}{2\pi\sigma^2} \exp \left\{ -\frac{1}{2\sigma^2} ((\vec{a}, \vec{e})_v - \{f_v\} - \{\lambda\}(\vec{a}, \vec{e})_{v-1})^2 \right\},$$

при этом $\{\lambda\}$ является матрицей неизвестных коэффициентов размерностью V , $\{f_v\}$ – известные величины, зависящие от управляющего воздействия U .

σ^2 – матрица соответствующих дисперсий размером $V \times V$.

Такой подход является удобным средством для учета имеющихся качественных представлений о статистическом поведении наблюдаемых данных $S(n)$ и значений $(\vec{a}, \vec{e})$ при незнании детальных количественных характеристик, точно определяющих это описание. Именно такое сочетание наиболее характерно для большинства прикладных задач обработки РС. Качественные представления, основанные на физической сущности рассматриваемых задач, дают возможность задать распределения вероятностей для $S(n)$ и $(\vec{a}, \vec{e})$, причем если $S(n)$ или $(\vec{a}, \vec{e})$ представляют собой конечное множество, то в качестве неизвестных параметров рассматриваются вероятности этих значений. При обработке РС увеличение размерности приводит к снижению качества решения основной задачи как из-за сложности технической реализации алгоритмов обработки данных наблюдения $S(n)$, так и из-за утраты некоторой доли входной информации, которую неизбежно приходится затрачивать для определения значений или исключения неизвестных мешающих параметров. Поэтому разрабатываемую в условиях априорной неопределенности систему необходимо дополнять алгоритмом проверки правильности априорных предположений, положенных в основу принятого в данной задаче параметрического описания. При этом необходимо установить, верно ли качественно введенное описание (например, при фильтрации процесса $S(n)$ – соответствует ли действительности аппроксимация $S(n)$ полиномом данной степени с неизвестными коэффициентами или эта модель недостаточно адекватна) и достаточны ли мощности пространств представлений относительно $(\vec{a}, \vec{e})$. Такой алгоритм дает возможность при необходимости усложнить описание анализируемого процесса, увеличив число параметров или качественно изменив модель априорной неопределенности.

Представленный подход статистического описания $S(n)$ и $(\vec{a}, \vec{e})$ соответствует тому, что создается некоторая аналитическая модель для описания статистических свойств $S(n)$ и $(\vec{a}, \vec{e})$ с той степенью полноты и подробности, которая

соответствует имеющимся знаниям о закономерностях их поведения, физических свойствах, взаимосвязи между собой. Подобная модель является наиболее сжатым описанием имеющегося опыта, содержащего как результаты изучения данных закономерностей, так и эмпирические данные относительно $S(n)$ и $(\vec{a}, \vec{e})$.

На практике получить статистическое описание параметров $(\vec{a}, \vec{e})$ отдельного диктора можно лишь для реализации задачи аутентификации, в то время как для более широкого класса задач обработки РС такое представление достаточно проблематично, что связано с отсутствием априорного знания всех возможных голосов дикторов, подвергаемых обработке.

На практике кроме эмпирических данных всякая иная априорная информация относительно $S(n)$ и $(\vec{a}, \vec{e})$ отсутствует. Однако, если эти данные получены в обстановке, статистически однородной или хотя бы статистически связанной с той, в которой принимается решение по данным наблюдения $S(n)$, то они являются статистическим эквивалентом аналитических моделей для распределений вероятности $S(n)$ и $(\vec{a}, \vec{e})$, необходимых для нахождения оптимального правила принятия решения. Степень этой эквивалентности зависит от объема эмпирических данных, которые, в свою очередь, могут быть использованы по-разному: непосредственно для нахождения недостающих распределений вероятности; для оценки функциональной зависимости апостериорного риска от $S(n)$ и решения U ; для уточнения структуры и параметров системы обработки РС, т. е. алгоритма ее функционирования.

Для анализа качественного состава имеющихся эмпирических данных и способов их использования при разработке систем в условиях априорной неопределенности предполагается факт наличия повторяющихся ситуаций, статистически однородных связанных с той, для которой принимается решение. Необходима оценка значений $(\vec{a}, \vec{e})$, позволяющая определить величину потерь $g[u, (\vec{a}, \vec{e}), S(n)]$ при принятии решения U по данным наблюдения $S(n)$, на основе которых принимается решение. В случае наличия N ситуаций, сходных или связанных с основной, для которых получены эмпирические зависимости, каждая из них для $v = 1, 2, \dots, N$

характеризуется параметрами, имеющими ту же природу, состав и содержание, что и параметры $(\vec{a}, \vec{e})$.

Таким образом, если принимается решение, $(\vec{a}, \vec{e}) \in \Lambda$, где Λ – множество значений параметров $(\vec{a}, \vec{e})$, то и каждое из значений является одним из элементов множества Λ ; если $(\vec{a}, \vec{e})$ имеет распределение с плотностью $p(\vec{a}, \vec{e})$, то и каждое из значений $(\vec{a}, \vec{e})$ имеет такое же распределение.

В общем случае эмпирические данные представляют собой совокупность значений $S_1(n), S_2(n), \dots, S_\Lambda(n)$, а каждая из составляющих $S_v(n)$ этой совокупности соответствует значению $(\vec{a}, \vec{e})_v$. Характер, значимость и способы использования этих сведений существенно различаются в зависимости от того, производится ли в v -й ситуации только наблюдение и регистрация каких-либо данных (аналогичных $S(n)$ для основной рабочей ситуации, либо как-то отличающихся от $S(n)$), или наряду с наблюдением, как и в рабочей ситуации, осуществляется принятие решения u_v .

При обработке речевых данных для всех $v = 1, \dots, N$ в рабочей ситуации решения $u_1, \dots, u_\Lambda$ и u выбираются независимо, причем решение u_v выбирается из условия минимума ожидаемых потерь для функции $g(u_v, (\vec{a}, \vec{e})_v)$, а решение u – для $g(u, (\vec{a}, \vec{e}))$. Решения $g(u_v, (\vec{a}, \vec{e})_v)$, могут быть как одинаковыми $g(u, (\vec{a}, \vec{e}))$, так и различными. Независимый выбор решений означает требование локальной оптимальности для каждого элемента серии $v = 1, \dots, N$. Любая из составляющих u_v наряду с данными $S(n)_v$, подобными $S(n)$, может содержать сведения о $(\vec{a}, \vec{e})$, u_v и g_v .

Довольно часто совокупность $\{y_1, \dots, y_N\}$ представляет собой упорядоченную последовательность. При этом каждое очередное решение u_v может зависеть от всех $y_1, y_2, \dots, y_N$. Если последовательность решений $u_1, \dots, u_N$ объединена единой целью, количественным выражением которого является функция потерь $g(u, u_1, u_2, \dots, u_N, \lambda, \lambda_1, \dots, \lambda_N)$, то требуется минимизировать общий риск, соответствующий этой функции потерь. Решения $u_1, \dots, u_N$ и u уже не могут приниматься

независимо, а выбор частных решений u_v в процессе набора эмпирических данных должен обеспечивать как уменьшение риска в каждой v -й ситуации, так и его минимальное значение в целом. При этом последовательность решений $u_1, u_2, \dots, u_N, u$ представляет собой многошаговый процесс.

Два наиболее типичного случая задание меры потерь соответствует аддитивной функции потерь, когда справедливо равенство

$$g(u, u_1, u_2, \dots, u_N, \lambda, \lambda_1, \dots, \lambda_N) = g(u, \lambda) + \sum_{v=1}^N g_v(u_v, \lambda_v),$$

а функции потерь не зависит от $\{u_v, \lambda_v\}$ т. е.

$$g(u, u_1, u_2, \dots, u_N, \lambda, \lambda_1, \dots, \lambda_N) = g(u, \lambda).$$

Последнее выражение определяет выбор решений u_v в процессе набора эмпирических данных для обеспечения минимальных потерь. Для функции потерь – зависимость в выборе решений u_v будет только тогда, когда эмпирические данные действительно имеют ценность, т. е. существует априорная неопределенность, связанная с неполным знанием функции правдоподобия $p(S(n)|\lambda)$ и априорного распределения $p(\lambda)$, или имеется статистическая зависимость между $S(n)$ и $y_1, \dots, y_N$. В противном случае многошаговый процесс принятия решений распадается на отдельные независимые шаги, и выбор последовательности решений $\{u_v\}$ не влияет на выбор решения u в целом. Для случая обработки РС функция потерь, по сути, вырождается в классическое ОСШ или субъективную оценку восприятия РС.

При наличии эмпирических данных может быть различная степень априорной неопределенности в отношении статистического описания $S(n)$, λ и $\{y_1, \dots, y_N\}$ от полного незнания их распределений вероятности до полного статистического описания.

При этом оптимальное правило решения $u = U(S(n))$ оказывается зависящим не от всей совокупности наблюдаемых данных $S(n)$ непосредственно, которые мо-

гут иметь очень большую или даже неограниченную размерность, а от сравнительно небольшой совокупности величин, связанных с наблюдаемыми данными $S(n)$.

При обработке РС для формирования управляющего воздействия при многоальтернативном решении вводится $m - 1$ функций $S(n)$, например функции вида

$$\Lambda_{j,m} = P_j(S(n)) / P_m(S(n)), j = 1, 2, \dots, m - 1,$$

которые отражают всю имеющуюся в $S(n)$ информацию, необходимую для принятия решения, поскольку оптимальное решение при любом $S(n)$ может быть выражено только через эти функции.

Подобным же свойством обладают значения $(\hat{\vec{a}}, \hat{\vec{e}})$, для которых достигается максимум апостериорной плотности вероятности при оценке параметров $(\vec{a}, \vec{e})$. В данном случае $(\hat{\vec{a}}, \hat{\vec{e}})$ является оптимальной оценкой вектора параметров $(\vec{a}, \vec{e})$.

На практике можно вообще не знать значение $S(n)$, а знать только значения некоторых функций от $S(n)$ - $\Lambda_{12}(S(n))$, $\lambda_{jm}(S(n))$, $(\hat{\vec{a}}, \hat{\vec{e}})$ соответственно, что и используется при низкоскоростном кодировании РС.

Таким образом, существуют преобразования, содержащие всю необходимую для принятия решения информацию, получаемую при наблюдении. Такие преобразования называются *достаточными статистиками* и играют важнейшую роль при нахождении как байесовых, так и небайесовых оптимальных правил принятия решения, т. е. при разработке оптимальных информационных систем [41]. Достаточная статистика тем более интересна, чем большее сжатие входных данных $S(n)$ она обеспечивает, т. е. чем меньше ее размерность. Преобразование минимальной размерности называется минимальной достаточной статистикой для данной задачи или данного класса задач.

Для каждой конкретной задачи обработки РС при определенных функциях потерь и априорном распределении минимальной достаточной статистикой является сама решающая функция $u = u(S(n))$ или любое взаимно однозначное преобра-

зование функции $u(S(n))$. Наибольшую ценность представляют достаточные статистики, минимальные для целого класса задач и позволяющие воспользоваться ими для решения всех задач этого класса. Так, преобразование вида $z = \{\Lambda_{1m}(S(n)), \dots, \Lambda_{m-1,m}(S(n))\}$, где $\Lambda_{jm}(S(n))$ – определяемая минимальная достаточная статистика для всей совокупности многоальтернативных задач; положение максимума апостериорной плотности вероятности $\hat{\lambda}(x)$ – это минимальная достаточная статистика для всех задач оценки с симметричной относительно разности $\lambda - \hat{\lambda}(x)$ апостериорной плотностью вероятности и относительно разности $u - \lambda$ функцией потерь.

При обработке РС наиболее универсальный способ отыскания достаточных статистик малой размерности заключается в анализе функции правдоподобия $P(S(n)|(\vec{a}, \vec{e}))$. Преобразование вида $\xi = \xi(S(n)) = \{Z(S(n)), \zeta(S(n))\}$, где размерность $z = z(S(n))$ меньше размерности $S(n)$, а функция $\zeta(S(n))$ дополняет преобразование $z(S(n))$ до взаимно однозначного. В силу взаимной однозначности $S(n)$ и ξ используется для описания полной совокупности данных наблюдения не $S(n)$, а $\xi = \xi(S(n))$. При этом решение u будет функцией $\xi = \{z, \zeta\}$ и тогда вместо функции правдоподобия $P(S(n)|(\vec{a}, \vec{e}))$ нужно рассматривать равенство

$$P(\xi|(\vec{a}, \vec{e})) = P(z, \zeta|(\vec{a}, \vec{e})) = P(\zeta|z, (\vec{a}, \vec{e}))P_z(z|(\vec{a}, \vec{e})).$$

Если z и ζ выбраны так, что распределение вероятности ζ не зависит от $(\vec{a}, \vec{e})$, т. е. $P(\zeta|z, \lambda) = P(\zeta, z)$, то преобразование $z = z(x)$ является достаточной статистикой. Преобразование $\zeta = \zeta(S(n))$ описывает ту часть сведений, содержащихся в совокупности входных данных $S(n)$, которая не несет никакой информации о λ и, следовательно, о последствиях от принятия того или иного решения и не участвует при его формировании.

Если преобразование $z = z(S(n))$, удовлетворяя приведенным выше условиям, имеет наименьшую размерность, то оно является минимальной достаточной статисти-

стикой для всего класса задач, связанных с принятием любых решений, последствия которых зависят от n , какие бы ни вводились функции потерь и априорные распределения вероятности. Дальнейшее сжатие достаточной статистики для конкретной задачи обработки РС возможно только при дополнительных ограничениях функции потерь и априорных распределений или при конкретном их задании.

Фактически нахождение правила решения $u(s(n))$, удовлетворяющего одному или сразу нескольким рассмотренным выше принципам предпочтения, представляет собой непростую задачу. Это связано с тем, что даже задача отыскания байесова правила решения при отсутствии априорной неопределенности, несмотря на существование общих рецептов, оказывается слишком сложной из-за трудностей вычисления апостериорного риска в явной и вполне детализированной форме. В этом отношении большую пользу при решении конкретных задач приносит использование достаточных статистик (минимальных или хотя бы достаточно малой размерности).

Для практических задач исследования систем обработки РС характерно наличие большой размерности – в первую очередь, данных $S(n)$, описывающих результаты наблюдения, что существенно затрудняет получение окончательных результатов. Использование достаточных статистик резко понижает размерность задачи, что приносит упрощение и при чисто байесовом подходе, а в задачах с априорной неопределенностью часто позволяет получить конкретные алгоритмы принятия решений, основанные только на достаточных статистиках и являющиеся если не абсолютно оптимальными, то, по крайней мере, обладающие многими чертами оптимальности. Естественным, что при относительно малой размерности легче выявить условия существования строго или приближенно равномерно наилучшего правила решения.

Таким образом, можно утверждать, что выявление достаточных статистик малой размерности является если не обязательным, то целесообразным элементом решения задачи создания системы обработки в условиях априорной неопределенности.

Для такой задачи в условиях априорной неопределенности особое значение имеют инвариантные достаточные статистики, т. е. такие преобразования $z = z(S(n))$, распределения вероятности которых одинаковы при любом распределении вероятности из класса $(\bar{a}, \bar{e})$. Если инвариантная достаточная статистика существует, то можно ввести преобразование $\xi = \xi(S(n)) = \{z(S(n)), \zeta(S(n))\}$, при котором функция правдоподобия, записанная в новых переменных $\xi(z, \zeta)$ для всех элементов $P(\xi | (\bar{a}, \bar{e}))$ класса $T_{(\bar{a}, \bar{e})}$, соответствующего имеющейся априорной неопределенности, может быть представлена в виде

$$P(\xi | (\bar{a}, \bar{e})) = P(z | (\bar{a}, \bar{e}))P(\zeta | z),$$

где $P(z | (\bar{a}, \bar{e}))$ – одинакова для всех $P(\xi | (\bar{a}, \bar{e})) \in T_{(\bar{a}, \bar{e})}$.

Отсюда следует, что апостериорное распределение $(\bar{a}, \bar{e})$ одинаково для всех элементов $T_{(\bar{a}, \bar{e})}$ и, следовательно, априорная неопределенность является несущественной. Отыскание инвариантной достаточной статистики фактически эквивалентно устранению первоначально имевшейся априорной неопределенности.

Тем не менее, общие методы отыскания инвариантных или приближенно инвариантных достаточных статистик для задач обработки РС не разработаны, а существование строго инвариантных достаточных статистик – довольно редкое явление. Фактически подобные задачи отражают широкий круг практических потребностей, связанных с нахождением правил проверки какого-либо однозначно или многозначно заданного гипотетического предположения, когда альтернативой этому предположению может быть все, что угодно.

Аналогичными свойствами строгой или приближенной инвариантности и достаточности обладают статистики, формируемые при использовании известных критериев согласия Колмогорова, Смирнова, Мизеса, хи-квадрат и т. д. [87, 104], применяемых для проверки гипотезы о соответствии данных наблюдения заданному распределению вероятности либо о его неизменности для двух совокупностей наблюдаемых данных.

При обработке РС минимаксный подход является удобным, а иногда и единственным средством получения правила решения $u(S(n))$ в тех случаях, когда априорная неопределенность распространяется только на распределение вероятности параметров $(\vec{a}, \vec{e})$, не наблюдаемых непосредственно, но влияющих на последствия от принятия решения. Особенно велико его значение для ситуаций, когда данные наблюдения $S(n)$ не содержат сведений об априорном распределении $(\vec{a}, \vec{e})$. Это имеет место в том случае, если эти данные получены для единственного значения $(\vec{a}, \vec{e})$ или еще дополнительно для нескольких известных значений $(\vec{a}, \vec{e})$, но так, что относительная частота различных значений $(\vec{a}, \vec{e})$, не связана с распределением $p(\vec{a}, \vec{e})$.

Минимаксный подход имеет более широкое значение и может применяться и тогда, когда априорная неопределенность распространяется на статистическое описание и данных наблюдения $S(n)$ и параметров $(\vec{a}, \vec{e})$, однако имеет некоторую специфику, которая заслуживает отдельного рассмотрения.

Если имеется априорная неопределенность относится $(\vec{a}, \vec{e})$, когда априорное распределение $(\vec{a}, \vec{e})$ с плотностью $p(\vec{a}, \vec{e})$ полностью или частично неизвестно, то для всякого правила решения $u(S(n))$ средний риск будет принимать различные значения, соответствующие разным $p(\vec{a}, \vec{e})$ из множества T , содержащего все возможные при данном уровне априорной неопределенности распределения $(\vec{a}, \vec{e})$. Если данные наблюдения $S(n)$ не содержат сведений о распределении $(\vec{a}, \vec{e})$, то класс минимаксных правил решений может быть сужен до множества $U(S(n))$ всех байесовых решений, соответствующих возможным априорным плотностям вероятности $p(\vec{a}, \vec{e}) \in T$. Это утверждение следует из теорем о полноте байесова класса решений [98] и условий их справедливости, которые в данном случае выполняются.

При этом также имеет место равенство минимакса и максимина для риска $R(u(S(n)), p)$, т. е.

$$\max_{(p \in T_0)} \min_{(u(S(n)))} R(u(S(n)), p) = \min_{(u(S(n)))} \max_{(p \in T_0)} R(u(S(n)), p),$$

откуда следует, что отыскание минимаксного правила решения можно свести к процедуре отыскания наименее предпочтительного распределения вероятности с плотностью $p(S(n)) \in T_0$ и последующему нахождению байесова правила решения относительно этого априорного распределения.

Следует отметить, что минимаксная задача может решаться не относительно общих мер, заданных на множестве T_0 возможных распределений вероятности $\{P(S(n)|(\bar{a}, \bar{e}))\}$, а непосредственно относительно $p(\bar{a}, \bar{e})$, заданных на множестве возможных значений $(\bar{a}, \bar{e})$.

При обработке речевых сигналов также возникают случаи, когда какие-либо сведения о статистических свойствах $(\bar{a}, \bar{e})$ отсутствуют вообще. При этом T множество любых неотрицательных функций, подчиненных условию нормировки на множестве возможных значений $(\bar{a}, \bar{e})$.

Для нахождения минимаксного правила полной априорной неопределенности относительно $(\bar{a}, \bar{e})$ используется теорема А. Вальда [72], согласно которой минимаксное правило решения $u_m(S(n))$ является байесовым относительно некоторого наименее предпочтительного распределения вероятности с плотностью $p^*(\bar{a}, \bar{e})$ и обладает тем свойством, что величина условного риска $r(u_m(S(n)), (\bar{a}, \bar{e}))$ для этого решения одинакова при всех значениях $(\bar{a}, \bar{e})$, для которых плотность $p^*(\bar{a}, \bar{e})$ отлична от нуля.

Минимаксный подход может быть использован для решения очень большого числа задач обработки РС, в которых имеет место априорная неопределенность в статистическом описании данных наблюдения $S(n)$. Более того, если в какой-то из этих задач существует равномерно наилучшее правило решения, то оно обязательно будет найдено при исследовании семейства минимаксных решений. Однако, нельзя указать какой-либо достаточно общей и эффективной процедуры нахождения минимаксных правил решения. Фактически нахождение минимаксного правила сводится к эмпирическому нахождению структуры решения. Таким

образом, если выбранное в условиях априорной неопределенности правило решения удовлетворяет принципу минимакса, то это гарантирует некоторый уровень риска, понизить который при функционировании системы обработки практически невозможно.

В условиях недостаточной априорной информации относительно свойств источника речевых данных имеет смысл использовать адаптивный байесов подход, основная идея которого заключается в следующем.

Пусть имеется априорная неопределенность в статистическом описании данных наблюдения $S(n)$ параметров $(\vec{a}, \vec{e})$, влияющих на последствия принимаемых решений или того и другого. Эта неопределенность не позволяет использовать классический подход: найти для любого возможного правила решения величину среднего риска $R(u(S(n)))$, величину апостериорного риска $R(u, S(n))$ и определить минимум апостериорного риска по u , т. е. найти значение $u = u_0(S(n))$, минимизирующее апостериорный риск и представляющее собой оптимальное байесово решение. Практически для любых конкретных условий в пределах имеющейся априорной неопределенности существуют вполне определенные, но неизвестные истинные значения величин $R[u(S(n))] = R_{\text{ист}}[u(S(n))]$ для всех $u(S(n))$, значения $R[u, S(n)] = R_{\text{ист}}[u, S(n)]$ для всех u и $S(n)$ и оптимальное решение $u_0(x)$, обращающее в минимум функцию $R_{\text{ист}}[u, S(n)]$.

Можно также использовать порядок, принятый при нахождении байесова правила решения, главным элементом которого является нахождение минимума апостериорного риска по u и положения этого минимума. Для этого необходимо использовать часть сведений, содержащихся в совокупности данных наблюдения $S(n)$, для оценки истинной величины апостериорного риска с тем, чтобы получить его приближенное значение для всех возможных решений u или хотя бы в окрестности его минимума. После того, как такая оценка получена, остается только воспользоваться стандартными рецептами для нахождения байесова решения, минимизировав вместо истинного оценочное значение апостериорного риска.

Если оставшаяся после оценки апостериорного риска часть сведений из общей совокупности данных наблюдения $S(n)$ достаточна для получения минимума оценочного значения апостериорного риска по u для значения, неодинакового при разных $S_v(n)$, то получено решение задачи разработки системы в условиях априорной неопределенности. Если же использованная оценка апостериорного риска является состоятельной и близка к его истинному значению, то найденное таким способом решение будет достаточно близко к оптимальному байесову решению при отсутствии априорной неопределенности, когда статистические описания $[S, (\vec{a}, \vec{e})]$ и истинные значения среднего и апостериорного риска для всех u известны полностью.

Такой подход позволяет получить в условиях априорной неопределенности правила решения (алгоритмы обработки информации), которые, даже при сравнительно небольшом количестве «избыточных» данных, позволяют найти правила решения, эффективность которых близка к соответствующим оптимальным байесовым правилам. При этом увеличение объема полной совокупности данных наблюдения $S(n)$ дает возможность осуществить достаточно быструю сходимость к результатам, которые имели бы место при отсутствии априорной неопределенности.

Таким образом, необходимо в условиях априорной неопределенности, характерной для задач обработки РС, сохранить для выбора наилучшего решения и нахождения наилучшего правила решения (алгоритма обработки информации) основное содержание байесова подхода. Оно заключается в том, что нужно сопоставлять каждому состоянию системы обработки данные наблюдений $S(n)$ и ожидаемое значение потерь, после чего выбрать решения исходя из минимума ожидаемых при данном состоянии знаний потерь, и считать оптимальным такое правило решения $u(S(n))$, которое соответствует минимуму ожидаемых потерь для данного значения $S(n)$. Именно на этом принципе осуществляется функционирование системы обработки РС, а задание функции потерь, функции правдоподобия, априорного распре-

деления, правила вычисления апостериорного распределения вероятности и апостериорного риска относятся больше к формальной и излишне формализованной, а не смысловой и практической стороне подхода.

Наилучшим решением является то, которое обеспечивает наименьшее значение ожидаемых потерь, что вовсе не зависит от объема имеющихся априорных данных. Степень полноты этих данных влияет на качество оценки ожидаемых потерь. При отсутствии априорной неопределенности ожидаемые потери для каждого из возможных решений u имеют точную количественную меру – величину апостериорного риска, которая определяется по формальным правилам как математическое ожидание потерь при данном состоянии знаний об обрабатываемом сигнале. Такая оценка ожидаемых потерь (расчет величины апостериорного риска) производится по данным наблюдения $S(n)$, а роль имеющихся априорных сведений о виде функции потерь и распределений вероятности $S(n)$ и $(\vec{a}, \vec{e})$ заключается только в том, что они позволяют установить четкое функциональное соответствие между значением $S(n)$ и величиной ожидаемых при принятии решения u потерь.

При наличии априорной неопределенности необходимо оценить ожидаемые потери для каждого возможного решения. Как и при отсутствии априорной неопределенности, эта оценка делается по имеющимся данным наблюдения $S(n)$. Однако из-за недостатка априорных сведений невозможно установить точное функциональное соответствие между $S(n)$, $(\vec{a}, \vec{e})$ и величиной апостериорного риска. В связи с этим роль и значение получаемой информации о данных наблюдения $S(n)$ повышаются, и эти данные выступают не просто в качестве аргумента известной функции апостериорного риска $R[u, S(n)]$, оценивающей ожидаемые потери, но и должны использоваться для восстановления функционального соответствия между ними и величиной ожидаемых потерь.

Процесс восстановления этого соответствия, обеспечивающий реализуемость основного принципа байесова подхода – выбора решения по минимуму ожидаемых потерь, можно назвать адаптацией, а соответствующие правила решений u

, устанавливающие функциональное соответствие между $S(n)$ и решением u , минимизирующим полученную с учетом ограниченности знаний оценку ожидаемых потерь, – адаптивными байесовыми правилами решения.

Такой подход основан на выборе решения по минимуму ожидаемых потерь и не отличается от обычного байесова подхода, обеспечивающего нахождение наилучших правил решения. Тем не менее, этот подход основан на замене точной меры ожидаемых потерь – апостериорного риска – его состоятельной оценкой по имеющимся данным наблюдения $S(n)$ и последующем применении всех формальных процедур перехода от заданной функции апостериорного риска к оптимальному правилу решения. Состоятельность оценки апостериорного риска позволяет быть уверенным в том, что не будут допущены грубые ошибки при сопоставлении величин ожидаемых потерь для разных решений и выборе наилучшего решения. Применение же формальных процедур нахождения байесова правила решения по величине апостериорного риска в задачах обработки РС дает гарантию, что ничего не ухудшится, и в дальнейшем получится наилучшее при данной оценке ожидаемых потерь правило решения. Для того, чтобы это правило решения было наилучшим в условиях априорной неопределенности, нужно, чтобы и принятая оценка апостериорного риска была наилучшей, а для того, чтобы правило решения было близко к наилучшему при отсутствии априорной неопределенности, нужно, чтобы оценочное и истинное значения апостериорного риска имели минимумы при одном и том же решении U_v почти для всех значений $S(n)$ при близких значениях U_v для всех значений $S(n)$.

При априорной неопределенности какие-либо сведения об аналитическом описании исходного материала полностью отсутствуют: неизвестно распределение вероятности наблюдаемых значений $x_v (v = 1, \dots, N + 1)$, полностью неизвестен вид функции потерь и природа и статистическое описание параметров $(\vec{a}, \vec{e})$, влияющих на величину потерь и последствия от принятия того или иного решения.

Таким образом, необходимо ввести ряд предположений о дискретности множества решений U , о дискретности множества значений S_v , о независимости и одинаковости распределений вероятности всех значений $S_v (v = 1, \dots, N + 1)$, об одинаковости истинных (неизвестных) функций потерь на всех шагах $v = 1, \dots, N + 1$ с требованием, чтобы полная совокупность данных наблюдения $S(n)$ содержала значения принятых при $v = 1, \dots, N$ решений u_v и появившихся при этом потерь g_v .

Указанные ограничения выражают разную форму представления имеющихся априорных знаний, отличную от параметрического статистического описания неизвестных распределений вероятности и функций потерь, а с учетом перечисленных ограничений необходимый для нахождения правила решения объем этих априорных знаний довольно велик.

Характерной чертой непараметрического случая является использование эмпирических распределений вероятности вместо истинных и эмпирических средних значений вместо математических ожиданий. Это обстоятельство приводит к определенным требованиям по объему и составу полной совокупности данных наблюдения $S(n)$, чтобы эмпирическое осреднение приводило к состоятельным оценкам необходимых для отыскания правил решения математических ожиданий (среднего риска, апостериорного риска, минимального значения апостериорного риска и т. д.). Указанная совокупность $\{S(n)\}$ должна иметь вполне определенный состав и содержать достаточное для построения таких оценок количество данных наблюдения.

Сформулируем следующее утверждение, лежащее в основе предлагаемого решения по непрерывной аутентификации диктора.

Утверждение 3. Нарушение характера и степени взаимосвязи элементов множеств декомпозиции речевого сигнала, определяемой многозначным отображением между соответствующими векторными множествами, указывает на наличие постороннего шума или несанкционированного использования системы кодирования РС.

Взаимосвязи множеств $(\vec{a}, \vec{e})$ в условиях акустических шумовых воздействий начинают разрушаться, что связано с неправильным выбором вектора $\vec{a}$ при поиске

в КК. В работе предложен и обоснован метод шумоподавления в РС на основе применения полиспектрального анализа. Возможность коррекции положения ЛСЧ за счет реализации технических решений по шумоподавлению представлена на (рисунках 5.9 – 5.16). Так, на основе анализа оптимизационной задачи создания адаптивной системы фильтрации РС, функционирующей в режиме реального масштаба времени обработки, разработаны методы шумоподавления на основе спектрального вычитания из-за их относительной простоты и возможности практической реализации на основе современных технологий ЦОС [11, 50, 141].

При организации доступа к ресурсу сети связи используются различные механизмы подтверждения легитимности абонента, основанные на биометрической аутентификации. Наиболее доступной является речевая аутентификация, которая позволяет гарантированно использовать ресурсы сети связи ограниченному кругу лиц. Основным недостатком подобных методов является однократная проверка легитимности пользователя с дальнейшим предоставлением ему прав использования ресурсов сети на весь сеанс связи. Для его устранения разработаны предложения по непрерывной аутентификации диктора, причем процедура аутентификации оказывается связанной непосредственно с процессом ведения телефонных переговоров по сети связи. В основе метода лежит использование низкоскоростного кодирования РС с применением ЛП с возбуждением от кода при использовании алгоритмов ВК параметров липредера и процедуры AbS.

Представленный подход, являющийся направлением дальнейших исследований, базируется на использовании зависимостей элементов декомпозиции РС, возникающих при ЛП с возбуждением от кода и ВК данных. Предлагаемые технические решения базируются на существующих технологиях обработки РС и могут быть использованы без изменения элементной базы построения подобных устройств. Как было отмечено (таблица 3.1), при оценке распределения информационных ресурсов в кадрах соответствующих липредеров можно заметить, что основная информационная составляющая передачи расходуется на представление сигнала возбуждения. Низкоскоростные липредеры являются архиваторами с высоким уровнем уплотнения информации. Идентификация сигнала возбуждения для

них является примерно в 4 раза более ресурсоемкой задачей по сравнению с простой задачей идентификации и передачи параметров модели, при этом, чем выше коэффициент сжатия речевой информации, тем больше внимания уделяется идентификации сигнала возбуждения и формированию его архива в виде векторов КК. Для уменьшения информационной избыточности представления данных при их передаче по КС используются статистические зависимости между элементами декомпозиции РС при его обработке на основе алгоритмов CELP и кластерного анализа данных. ВК позволяет эффективно кодировать РС с потерями при использовании КК. Таким образом, КК представляет собой хранилище опорных векторов, каждый из которых имеет порядковый номер. С каждым из векторов КК параметров формирующей системы голосового тракта связана область кластерного подпространства КК векторов сигналов возбуждения. После процедуры обучения, в результате которой определяются данные подпространства для всех векторов КК параметров формирующей системы голосового тракта, осуществляется запись в память системы обработки соответствующих зависимостей.

При функционировании системы обработки РС вычисленный вектор КК определяет подпространство векторов сигналов возбуждения синтезирующей модели. Использование такой реструктуризации КК позволяет значительно сократить среднюю скорость передачи данных в КС. Также такой подход дает возможность учитывать индивидуальные особенности диктора и осуществлять процедуру непрерывной речевой аутентификации в процессе функционирования системы обработки РС [12, 13]. На (рисунке 3.7) представлена упрощенная схема аутентификации легитимного диктора. Вычисленный вектор параметров голосового тракта определяет индивидуальность диктора, отражая его формантные особенности, а их совокупность определяет динамику изменения формантных областей в процессе ведения разговора. По распределению значений нескольких подобных векторов можно сделать заключение о принадлежности диктора легитимной группе пользователей сети связи. При этом априорные вероятности появления объектов аутентификации с заданными свойствами являются неизвестными.

При использовании для низкоскоростного кодирования речи зависимостей между векторами КК, обученных на уникальной речи диктора, возникающие зависимости характеризуют их индивидуальные особенности, что связано с наличием структурно организованных связей между подпространствами КК векторов сигналов возбуждения и векторами КК параметров голосового тракта легитимного пользователя, которые будут нарушены при попытке использования системы связи.

Подобные зависимости взаимного отношения частот обращения к одинаковым областям КК для отдельного легитимного пользователя можно выделить, контролируя поток обращений к этим областям. Такой подход дает возможность использовать в абонентских устройствах обработки РС универсальные КК параметров декомпозиции. При этом нет необходимости в коренной модернизации уже работающего абонентского оборудования.

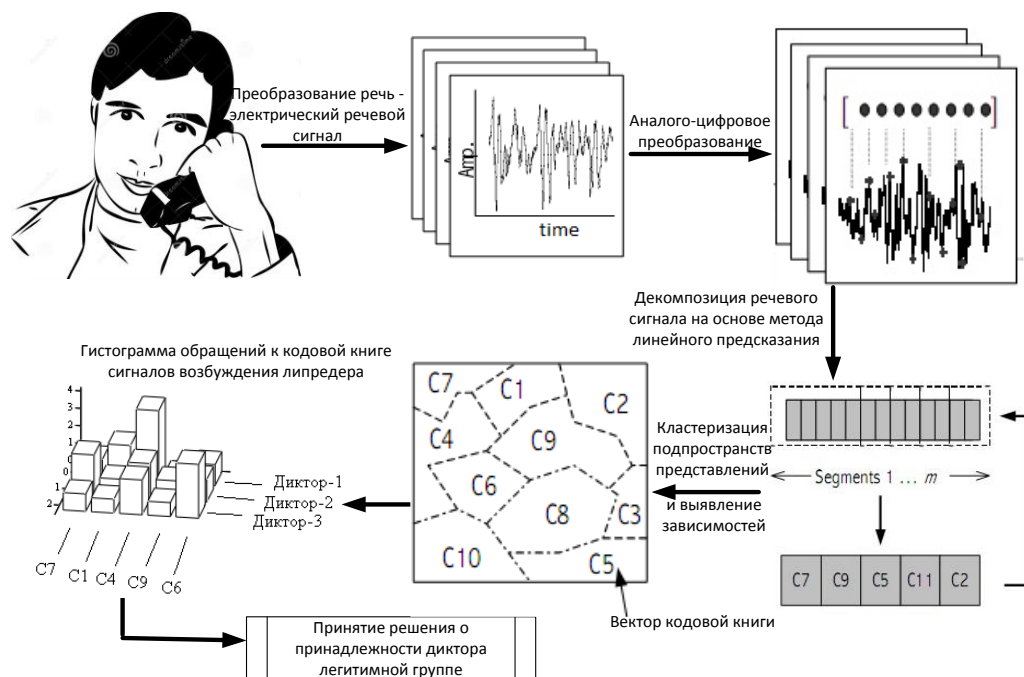


Рисунок 3.7 – Аутентификации легитимного диктора в сети связи

Поддержка решения о получении доступа к услугам сети связи с персонального абонентского терминала может осуществляться непрерывно. Если

терминал оказывается в "чужих" руках, то изменится поток обращения к областям КК. Изменение потока векторов КК для параметров "чужого" голоса приводит к иному выбору подпространства КК. Это разрушает выявленные статистические связи между элементами декомпозиции РС, что позволяет выявить факт несанкционированного доступа к сети связи. Применение зависимости элементов декомпозиции РС при ЛП на основе CELP-подобных алгоритмов в процессе его обработки и низкоскоростного кодирования позволяет по-новому рассмотреть задачу создания индивидуально ориентированных систем абонентского речевого кодирования. Подобная система обработки РС позволяет достигать значительного коэффициента сжатия по сравнению с существующими вокодерами с ЛП. При этом индивидуализация абонентского устройства может осуществляться в процессе его использования путем поэтапного создания индивидуальных КК элементов декомпозиции РС на речи пользователя в результате самообучения системы, использующей стандартный и полностью автоматический алгоритм обучения больших искусственных нейронных сетей по ГОСТ Р 52633.5.

Следует подчеркнуть, что наличие полностью автоматизированного обучения – это принципиально важное требование к средствам биометрической аутентификации личности по базовому стандарту ГОСТ Р 52633.0.

Таким образом, использование многомерного статистического кластерного анализа, методов классического ЛП РС на основе CELP-моделей и предлагаемых технических решений позволяет одновременно со снижением средней скорости передачи в КС осуществить персонификацию абонентского доступа к сети связи.

К достоинствам представленного метода следует отнести тот факт, что уменьшение в кадре передачи количества бит определяющих положение векторов сигнала возбуждения позволяет снизить скорость передачи данных в КС, а также уменьшить вычислительную сложность алгоритма кодирования РС на передающей стороне. Возможно также понижение скорости передачи данных в КС по сравнению с известными способами кодирования, либо перераспределение информационного ресурса, предоставляемого КС, на формирование дополнительных сервисов абонентского обслуживания.

Использование представленных решений позволяет без ухудшения качества синтезированного сигнала сократить среднюю скорость передачи данных, а при использовании индивидуально ориентированных КК элементов декомпозиции РС существует возможность достигать и значительно более низких скоростей передачи, осуществляя постоянный контроль легитимности доступа к сети связи. Таким образом, при реализации предлагаемых решений было выявлено, что использование зависимостей параметров голосового тракта и сигналов возбуждения при разработке низкоскоростных систем кодирования РС дает возможность сокращения средней скорости передачи при сохранении разборчивости синтезированной речи, что показано в шестом разделе. Применение предлагаемого подхода возможно в системах обработки РС, использующих в своей основе процедуру ЛП и кластерное представление параметров липредера.

Сигнал возбуждения формирующего фильтра в общем идеальном случае рассматривается как результат аппроксимации ошибки ЛП. При его формировании в качестве критерия используется минимум дисперсии ошибки оценивания. Применение оценки ФПВ исследуемых процессов и критерия разработки системы на их основе слабо влияет на улучшение слухового восприятия, что связано с отсутствием в ФПВ данных об особенностях формирования анализируемой последовательности. Одной из вероятностных характеристик, использование которой дает положительные результаты при формировании порождающего процесса, является анализ корреляционной функции. В теории ЛП порождающий процесс должен быть некоррелированным, однако на практике ввиду ограниченности порядка формирующей модели данное требование не выполняется.

Реализуемая степень оптимизации во многом определяется уровнем имеющейся априорной информации об обрабатываемых параметрах и степенью ее алгоритмического использования. Очевидно, что применение локально-стационарной модели РС не устраняет существующую априорную неопределенность относительно статистических и параметрических характеристик речевого источника с памятью. Последовательность $u(n)$ в (3.3) соответствует белому шумовому процессу,

который в теории используется в качестве входного воздействия формирующего фильтра. Однако последовательность значений ошибки ЛП, используемая для возбуждения, будет взаимосвязана с оценкой ЛП (предсказанным сигналом) до тех пор, пока не будет представлять собой белый шумовой процесс, чего на самом деле не произойдет ввиду конечности порядка M . В [] также отражена взаимосвязь значений параметров голосового тракта с возбуждающим сигналом, что явилось предпосылкой для сокращения избыточности представления РС при его обработке путем локализации подпространства представления сигналов возбуждения. Определение параметров возбуждающего сигнала в системе анализа с ЛП обычно основано на исследовании сигнала ошибки $e(n)$. Полученный сигнал возбуждения формирующего фильтра голосового тракта является аппроксимацией сигнала ошибки предсказания. На основании данного сигнала путем объединения подпространств представления зависимых сигналов с их последующей обработкой формируется КК векторов возбуждения.

Таким образом, используя взаимосвязь между вектором сигнала возбуждения фильтра синтеза и вектором параметров голосового тракта можно понизить скорость передачи при низкоскоростной обработке РС в условиях наличия акустических шумов, не передавая полную информацию об одной из составляющих, обычно входящих в кадр передачи [51, 58].

3.5 Выводы по третьему разделу

1. Перспективным направлением развития систем обработки РС является создание систем, использующих зависимости между элементами декомпозиции сегментов РС, то есть систем, в которых анализ и обработка РС будет производиться на основе учета зависимостей параметров модели, представляющей голосовой тракт и сигнала возбуждения. Анализ современного уровня низкоскоростного кодирования РС на основе ЛП показывает, что в настоящее время данные элементы

обрабатываются независимо друг от друга, что требует дополнительного информационного ресурса на обработку, представление и описание параметров модели и сигналов возбуждения в кадре передачи.

2. Использование зависимостей элементов декомпозиции РС при ЛП с ВК его параметров позволяет усовершенствовать системы обработки РС, поскольку большинство подобных устройств осуществляли независимое представление и кодирование данных параметров, что определяло информационную избыточность их кадра передачи.

3. Учет взаимосвязей элементов декомпозиции позволяет уменьшить число бит, необходимых для представления векторов сигнала возбуждения, что приводит к снижению скорости передачи при сохранении качества синтезированной речи по показателям естественности ее звучания и узнаваемости.

4. Переменная скорость передачи реализуется выделением различного числа бит на кодирование параметров элементов декомпозиции РС при ЛП вследствие псевдослучайного характера следования в речи классифицированных сегментов и кластеров КК.

5. Использование зависимостей элементов декомпозиции РС при ЛП с ВК его параметров на КК, обученных на уникальной речи отдельного диктора позволяет разработать метод непрерывной аутентификации легитимного диктора.

6. На основе обобщенной математической модели кодирования РС разработана математическая модель системы низкоскоростного кодирования РС на основе метода линейного предсказания, отличающаяся от известных учётом зависимости между элементами его декомпозиции при ЛП. Разработанная математическая модель системы кодирования РС показывает пути совершенствования систем обработки РС в условиях соответствия тенденциям развития инфокоммуникационных систем при учете зависимостей элементов декомпозиции РС.

4. МЕТОД ШУМОПОДАВЛЕНИЯ В РЕЧЕВОМ СИГНАЛЕ НА ОСНОВЕ ПРИМЕНЕНИЯ ТЕХНОЛОГИЙ ПОЛИСПЕКТРАЛЬНОГО АНАЛИЗА

4.1 Вводные замечания

Разработка методов очистки РС от шума является весьма актуальным направлением исследований [216]. В тех случаях, когда шум имеет значительную интенсивность, его наличие может существенно исказить результаты обработки, анализа или распознавания РС. В целом ряде других случаев, например, при обработке сильно зашумленного РС или его восстановления на фоне шумов и помех, задача очистки РС от шума или его восстановления носит самостоятельный характер и является частной целью работы [228, 229].

В данном разделе развиты теоретические основы анализа и обработки речевого сигнала при шумоподавлении (подраздел 4.2), разработана математическая модель системы фильтрации РС на основе применением полиспектрального анализа (подраздел 4.3), представлены технические решения по выделению сегментов паузы и оценки характеристик акустических воздействий в условиях зашумления обрабатываемого РС (подраздел 4.4).

4.2 Обработка речевого сигнала при подавлении помех

На основе анализа методов фильтрации РС можно выделить следующие базовые подходы к вопросу фильтрации обрабатываемого зашумленного РС, такие, как винеровская фильтрации и методы, основанные на параметрическом анализе сигналов. Стоит отметить, что эффективность всех подходов зависит от ряда допущений и ограничений как к РС, так и к воздействию. Решение поставленной научной задачи, заключающейся в создании адаптивной системы фильтрации зашумленного РС, невозможно без учета ограничения на масштаб времени обработки [223]. Данный аспект не только накладывает свой отпечаток на решение вопроса моделирования системы фильтрации РС, но и выделяет методы шумоподавления на основе спектрального вычитания из-за их относительной простоты, что обуславливается, в свою очередь, возможностью практической реализации и современным

уровнем развития технологий ЦОС [184]. Инструменты методов на основе спектрального вычитания базируются на анализе в основном только мощностных характеристик акустического шумового воздействия, при этом существует сложность в адекватной оценке данных параметров, вследствие чего подобные классические методы не позволяют эффективно осуществлять фильтрацию РС в УИАВ [11, 50, 152, 184].

Метод на основе спектрального вычитания обобщенно представлен в виде блок-схемы (рисунок 4.1). Детализация процедур, показанных на (рисунок 4.1), представлена следующим образом:

- 1) Ввод непрерывного РС $U(t)$;
- 2) АЦП с частотой дискретизации $f_d = 8$ кГц;
- 3) Сегментация дискретного РС на участки локальной стационарности по 20 мс (160 отсчетов) $U(nT, d)$ при $n=0, 1, 2, \dots, N-1$; $d=1, 2, \dots, D$;
- 4) Оценка мгновенной энергии $O(d)$ и количества знакоперемен $\mu_s^\pm(d)$ на сегменте;
- 5) Выделение длительной M – сегментов паузы, так что:

$$D = L + M, \quad (4.1)$$

где D – общее количество сегментов обрабатываемого РС; M – количество сегментов паузы; L – количество активных сегментов.

- 6) Дискретное преобразование Фурье на сегменте при $n = 0, 1, 2, \dots, N-1$, $k = 0, 1, 2, \dots, N-1$:

$$\dot{A}_U(jk\Omega, d) = \sum_{n=0}^{N-1} U(nT, d) \cdot e^{-j \cdot \frac{2 \cdot \pi}{N} \cdot k \cdot n} = \sum_{n=0}^{N-1} U(nT, d) \cdot e^{-jk\Omega nT}, \quad (4.2)$$

$$\Omega = \frac{2 \cdot \pi}{N \cdot T}, \quad (4.3)$$

где $\dot{A}_U(jk\Omega)$ – комплексный спектр Фурье обрабатываемого сегмента зашумленного РС; k – порядковый номер спектральных составляющих дискретного сигнала; n – порядковый номер отсчетов дискретного сигнала; Ω – основная частота ДПФ.

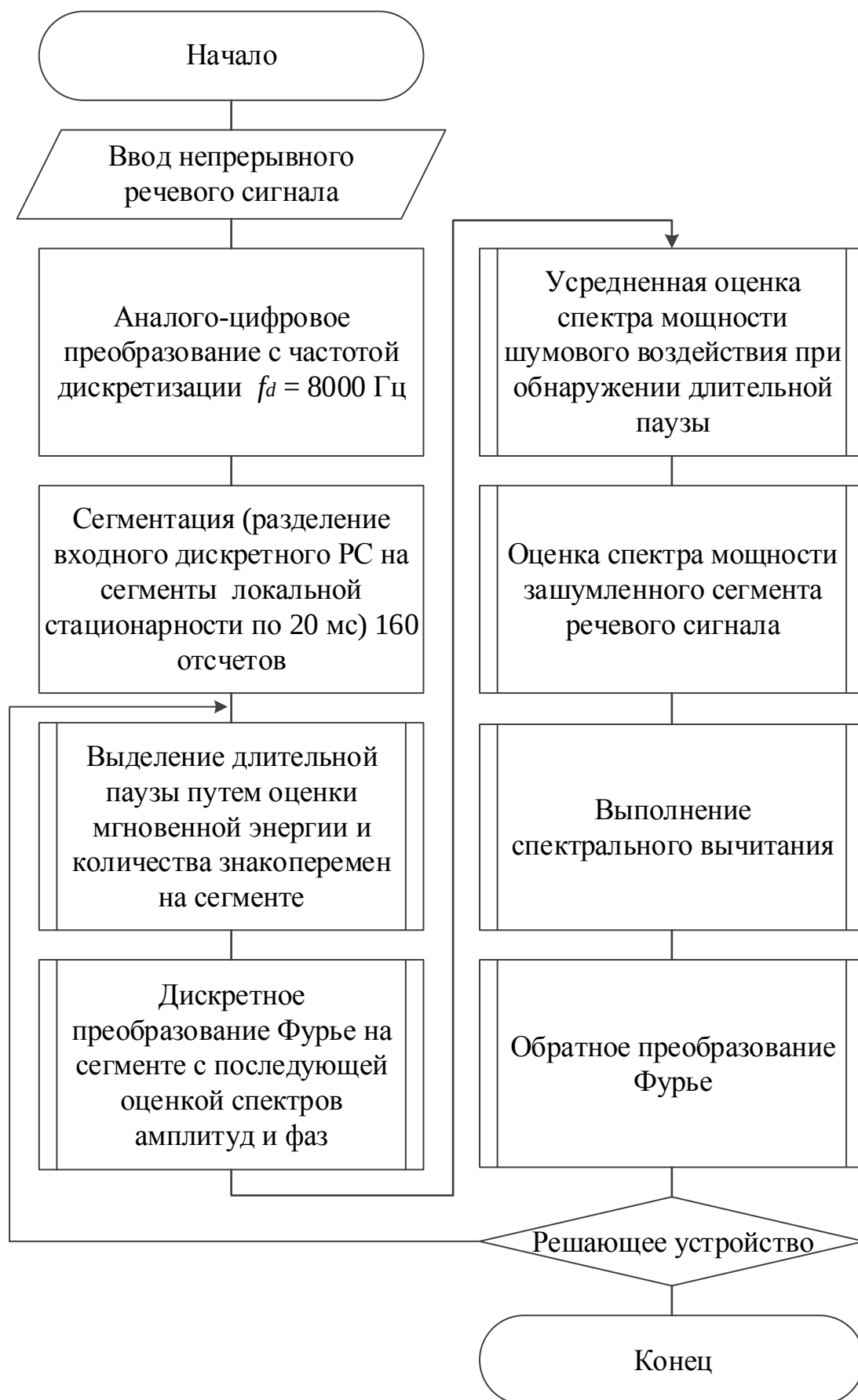


Рисунок 4.1 – Блок-схема алгоритма шумоподавления
на основе спектрального вычитания

7) Усредненная оценка спектра мощности воздействия по выражениям:

$$A_U(jk\Omega, d) = \sqrt[2]{\operatorname{Re}\left(\dot{A}_U(jk\Omega, d)\right)^2 + \operatorname{Im}\left(\dot{A}_U(jk\Omega, d)\right)^2} \quad (4.4)$$

$$A_U^2(jk\Omega, d) = \frac{1}{2} \cdot \left(\dot{A}_U(jk\Omega, d)\right)^2, \quad (4.5)$$

$$\left\langle A_E^2(jk\Omega, d) \right\rangle = \frac{1}{M} \cdot \sum_{m=1}^{m=M} A_U^2(jk\Omega, m), \quad (4.6)$$

где $\dot{A}_U(jk\Omega, d)$ – оценка комплексного спектра Фурье обрабатываемого сегмента РС; $A_U(jk\Omega, d)$ – оценка спектра амплитуд Фурье обрабатываемого сегмента РС; $A_U^2(jk\Omega, d)$ – оценка спектра мощности обрабатываемого сегмента РС; $\left\langle A_E^2(jk\Omega, d) \right\rangle$ – усредненная оценка спектра мощности воздействия по M сегментам паузы для сегмента d .

8) Оценка спектра мощности отфильтрованного сегмента d $A_{SS}^2(jk\Omega, d)$ выполнением спектрального вычитания при $k = 0, 1, 2, \dots, N-1$:

$$A_{SS}^2(jk\Omega, d) = \left| A_U^2(jk\Omega, d) - \left\langle A_E^2(jk\Omega, d) \right\rangle \right|, \quad (4.7)$$

где $A_{SS}(jk\Omega)$ – оценка спектра мощности отфильтрованного сегмента РС.

9) Оценка спектров амплитуд и фаз Фурье отфильтрованного сегмента d при $k=0, 1, 2, \dots, N-1$:

$$A_{SS}(jk\Omega, d) = \sqrt[2]{2 \cdot A_{SS}^2(jk\Omega, d)}, \quad (4.8)$$

$$F_U(jk\Omega, d) = \arctan \frac{\operatorname{Im}\left(\dot{A}_U(jk\Omega, d)\right)}{\operatorname{Re}\left(\dot{A}_U(jk\Omega, d)\right)}, \quad (4.9)$$

где $A_{SS}(jk\Omega, d)$ – оценка спектра амплитуд обработанного сегмента РС;
 $F_U(jk\Omega, d)$ – оценка спектра фаз обрабатываемого сегмента зашумленного РС.

10) Оценка комплексного спектра Фурье сигнала на выходе системы фильтрации для сегмента d и выполнение обратного дискретного преобразования Фурье при $n = 0, 1, 2, \dots, N-1, k = 0, 1, 2, \dots, N-1$:

$$\overset{\bullet}{A}_{SS}(jk\Omega, d) = A_{SS}(jk\Omega, d) \cdot \exp\left[F_U(jk\Omega, d)\right], \quad (4.10)$$

$$S^*(nT, d) = \frac{1}{N} \cdot \sum_{k=0}^{N-1} \overset{\bullet}{A}_{SS}(jk\Omega, d) \cdot e^{jk\Omega nT}, \quad (4.11)$$

где $\overset{\bullet}{A}_{SS}(jk\Omega, d)$ – оценка комплексного спектра Фурье сигнала на выходе системы фильтрации для сегмента d ; $S^*(nT, d)$ – последовательность отсчетов сегмента РС на выходе системы фильтрации.

В силу специфики ЦОС далее будут использованы следующие обозначения:
 d – порядковый номер сегмента локальной стационарности; $n=0, 1, 2, \dots, N-1$ – последовательный номер отсчета на сегменте локальной стационарности d (с учетом замены $nT \leftrightarrow n$); $k=0, 1, 2, \dots, N-1$ – порядковый номер спектральных составляющих дискретного сигнала.

На основе анализа выделены следующие недостатки методов с использованием спектрального вычитания:

- 1) необходимость обнаружения длительной паузы в обрабатываемом РС для оценки спектральных характеристик воздействия;
- 2) появление после спектрального вычитания в РС искажений, известных как "музыкальные тона";
- 3) пренебрежение влиянием зашумления на фазовый спектр сигнала;

Основным предположением, которое используется в алгоритмах шумоподавления (с одним входом) на основе спектрального вычитания, является то, что усредненная оценка спектра мощности паузы РС $\langle A_E^2(k, d) \rangle$ равна оценке аддитивного воздействия помехи на сам сигнал на зашумленном участке:

$$A_U^2(k, d) \approx A_{SS}^2(k, d) + \langle A_E^2(k, d) \rangle, \quad (4.12)$$

В свою очередь обработка РС при использовании локально-стационарной модели осуществляется на сегменте квазистационарности, что требует анализа характеристик акустических шумовых воздействий на коротких временных интервалах. На рисунке 4.2 представлены гистограммы последовательностей 256 отсчетов различных шумовых воздействий.

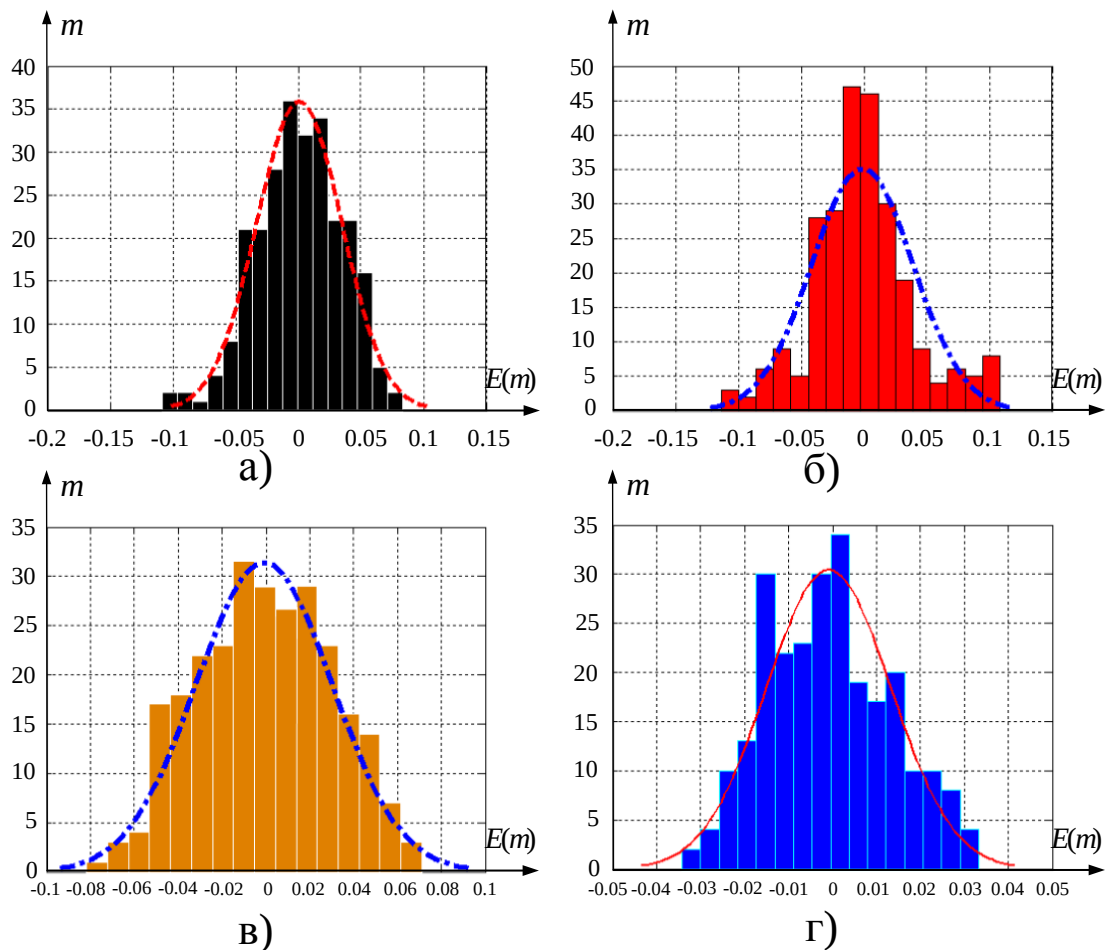


Рисунок 4.2 – Гистограммы мгновенных значений акустических шумовых сигналов на длительности 32 мс при частоте дискретизации 8 кГц (256 отсчетов):

а) шум автомагистрали; б) шум боя; в) шум поезда; г) шум города.

Исходя из анализа рисунка 4.2 можно сделать следующие выводы:

- 1) статистические характеристики акустических воздействий на малых по длительности интервалах наблюдения имеют отличия от «нормального» закона распределения, что указывает на необходимость изучения статистик более высокого порядка [70, 115];
- 2) реальные акустические воздействия *обладают высокой степенью симметричности* относительно своего математического ожидания;
- 3) как следствие первого и второго пункта вытекает возможность представления РС в виде суммы гауссовской и негауссовской компонент.
- 4) существует необходимость выявления параметров РС, которые зависят от факта посегментной обработки;
- 5) имеется потребность оценки интервала анализа, в пределах которого акустическое воздействие, с учетом различных допущений, сохраняет свои характеристики.

Проведенные исследования дали возможность выявить параметры, связанные с посегментной обработкой РС [11, 16, 18, 37, 44], которые далее используются в разработанном методе шумоподавления:

- 1) Коэффициент стабильности эффективной полосы частот акустического воздействия α , описывающий отношение максимального значения полосы частот $\Delta P(d)$ сегментов акустического воздействия, в которой сосредоточено 90% мощности, относительно своего среднего значения на квазистационарном длительном интервале наблюдения:

$$\alpha = \frac{\max[\Delta P(d)]}{\Delta P_a}, \quad (4.13)$$

где d – последовательный номер сегмента обработки в рамках интервала наблюдения при $d = 1, 2, 3, \dots, D$; ΔP_a – среднее значение эффективной полосы частот акустического воздействия на квазистационарном длительном интервале наблюдения, определяемое согласно следующего выражения:

$$\Delta P_a = \frac{1}{D} \cdot \sum_{d=1}^{d=D} \Delta P(d). \quad (4.14)$$

2) Коэффициент знакоперемен мгновенной энергии сегмента β , описывающий в физическом смысле частоту колебания энергии акустического воздействия с учетом посегментной обработки РС относительно своего среднего значения на квазистационарном длительном интервале наблюдения:

$$\beta = \frac{1}{D} \cdot \sum_{d=1}^{d=D} \mu^{\pm}(d), \quad (4.15)$$

где $\mu^{\pm}(d)$ – признак смены значения мгновенной энергии сегмента относительно среднего значения в рамках интервала наблюдения при $d = 1, 2, 3, \dots, D-1$, определяемый согласно следующего выражения:

$$\begin{aligned} O(d) < O_a \quad \& \quad Q(d+1) > O_a \quad \Rightarrow \quad \mu^{\pm}(d) = 1, \\ O(d) > O_a \quad \& \quad Q(d+1) < O_a \quad \Rightarrow \quad \mu^{\pm}(d) = 1, \\ O(d) \leq O_a \quad \& \quad O(d+1) \leq O_a \quad \Rightarrow \quad \mu^{\pm}(d) = 0, \\ O(d) \geq O_a \quad \& \quad O(d+1) \geq O_a \quad \Rightarrow \quad \mu^{\pm}(d) = 0, \end{aligned} \quad (4.16)$$

где $O(d)$ – мгновенная энергия сегмента d ; O_a – среднее значение мгновенной энергии сегмента в рамках интервала наблюдения при $d = 1, 2, 3, \dots, D$, определяемые согласно следующих выражений:

$$O(d) = \sum_{n=1}^{n=N} [E(n, d)]^2, \quad (4.17)$$

$$O_a = \frac{1}{D} \cdot \sum_{d=1}^{d=D} O(d), \quad (4.18)$$

где $E(n, d)$ – мгновенное значение отсчета n воздействия на сегменте d при $n = 1, 2, 3, \dots, N$; N – количество отсчетов на сегменте локальной стационарности.

Для практического применения важен также такой параметр акустического воздействия как пикфактор γ , определяемый как отношение максимального значения мгновенной энергии сегмента к ее среднему значению на длительном интервале наблюдения. Все результаты, полученные для анализа вышеописанных характеристик акустических воздействий, представлены в таблице 4.1. и проиллюстрированы на (рисунке 4.3, на примере шума ветра).

Таблица 4.1 – Параметрические характеристики шумовых воздействий*

Вид шума	Параметр											
	α				β				γ			
	1 сек	2 сек	4 сек	8 сек	1 сек	2 сек	4 сек	8 сек	1 сек	2 сек	4 сек	8 сек
КБГШ	1.16	1.18	1.15	1.13	0.54	0.46	0.51	0.47	1.15	1.29	1.31	1.38
Шум двигателя	1.86	1.89	1.86	1.96	0.36	0.40	0.41	0.37	1.91	2.09	2.12	2.09
Шум ветра	1.66	1.80	1.93	2.19	0.36	0.36	0.37	0.36	2.12	2.65	3.32	3.01
Шум города	1.81	1.83	1.84	1.84	0.40	0.39	0.37	0.36	1.90	2.68	2.39	3.01
Шум пламени	1.61	1.65	1.91	1.93	0.51	0.47	0.45	0.40	3.04	2.80	3.31	3.2
Шум боя	1.82	1.87	2.04	2.1	0.35	0.37	0.35	0.36	3.91	3.73	3.75	3.86
Шум вертолета	1.73	1.75	1.77	1.72	0.37	0.36	0.38	0.37	2.76	2.83	2.97	2.99
Шум поезда	1.67	1.70	1.68	1.70	0.32	0.35	0.37	0.33	2.07	2.10	2.12	2.12
Шум толпы	1.32	1.34	1.34	1.35	0.45	0.45	0.46	0.45	2.05	2.05	2.07	2.09
Шум автомагист.	1.23	1.24	1.25	1.27	0.44	0.43	0.46	0.45	2.13	2.18	2.23	2.34
Тоновой шум	1	1	1	1	0.51	0.48	0.52	0.51	1.17	1.18	1.17	1.16

* результаты, представленные в таблице, получены при частоте дискретизации 8000 Гц и сегменте 32 мс (256 отсчетов), имеют высокую коррелированность с испытаниями на других частотах дискретизации и других длительностях сегмента анализа.

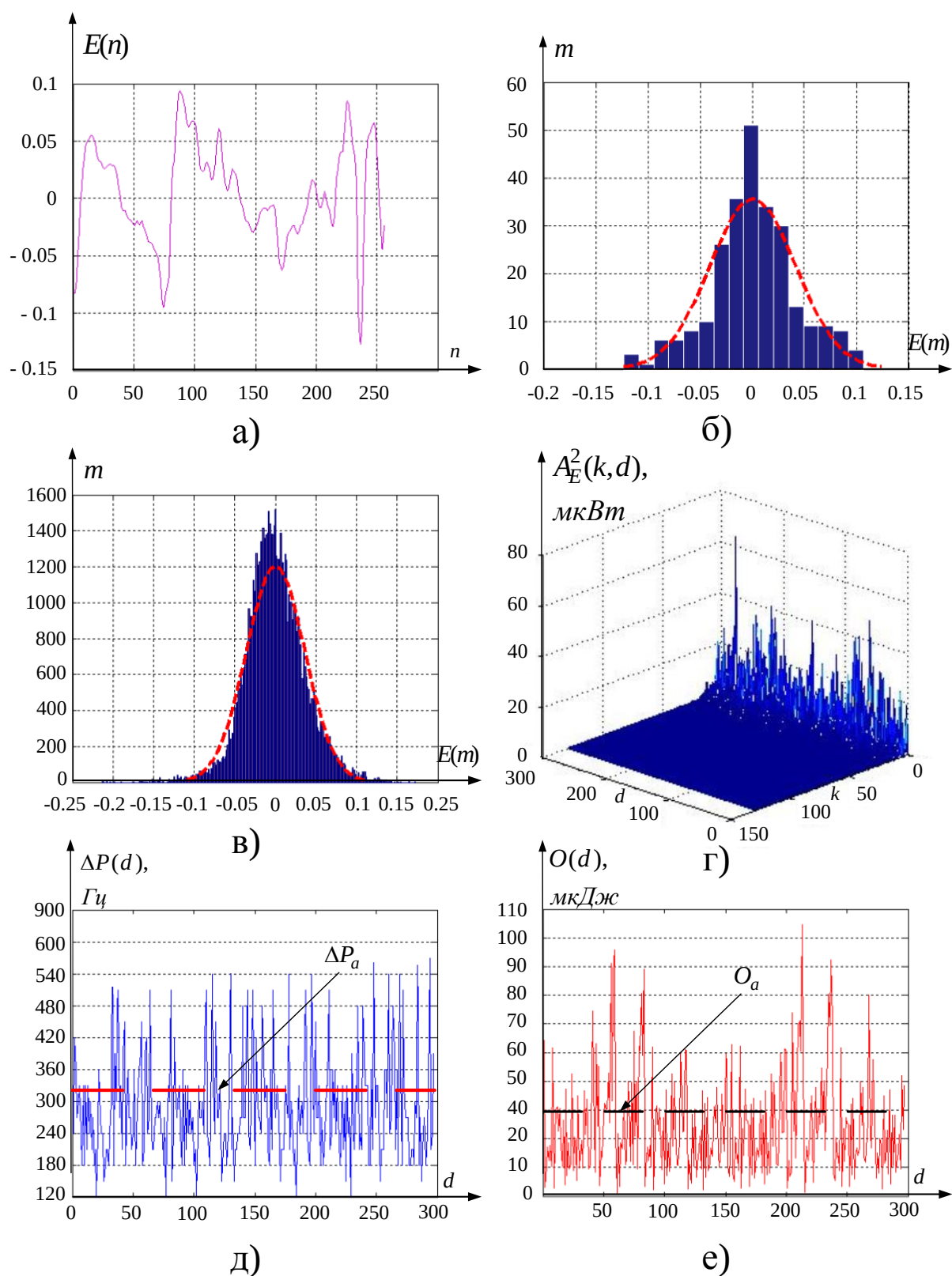


Рис. 4.3 – Параметры акустического воздействия: а) реализация на сегменте 32 мс (256 отсчетов); б) гистограмма мгновенных значений на сегменте; в) гистограмма мгновенных значений на интервале 10 секунд; г) спектр мощности посегментно на длительности 10 секунд; д) эффективная полоса частот посегментно; е) значение мгновенной энергии посегментно.

Анализ полученных результатов позволяет сделать следующие выводы:

1) Реальные акустические воздействия на сегменте локальной стационарности (рисунок 4.3, а) обладают низким количеством знакоперемен относительно амплитудной точки "0", при этом закон распределения мгновенных значений на сегменте (рисунок 4.3, б) имеет отличия от "нормального", однако обладает высокой степенью симметрии.

2) Закон распределения мгновенных значений стремится к гауссовскому характеру при увеличении количества анализируемых отсчетов РС (рисунок 4.3, в).

3) Значение мгновенной энергии сегмента акустических воздействий на длительном интервале имеет большое количество изменений относительно своего среднего (рисунок 4.3, г, е; таблица 4.1, параметр β).

4) Частотный состав и эффективная полоса акустического воздействия обладают высокой степенью стабильности относительно динамического диапазона в течение длительных интервалов наблюдения (рисунок 4.3 г, д; таблица 4.1 параметр α , сравнительный анализ параметров α и γ при одинаковых интервалах наблюдения для различных воздействий).

5) Интервал квазистационарности реальных акустических воздействий не более 2 секунд, то есть акустическое воздействие обладает высокой степенью стабильности своих параметров в течение 2 секунд.

Исходя из вышеизложенного можно сделать вывод о том, что большинство реальных акустических воздействий можно считать квазистационарными сигналами на длительных интервалах наблюдения (не более 2 секунд), в течение которых сохраняется стабильность статистических и параметрических характеристик.

В связи с этим сильно возрастает зависимость системы фильтрации на основе спектрального вычитания (с одним входом) от самого факта правильного обнаружения паузы. Результативность большинства методов обработки РС сильно зависит от эффективности применяемого алгоритма детектирования активности речи. Идеальный детектор активности речи, в рамках предметной области исследования, должен быть независимым от уровня акустического воздействия и быть зависимым

от максимума параметров приложения, в котором его используют. Применительно к вопросу оценки характеристик воздействия можно выдвинуть следующие требования к методам выделения паузы в обрабатываемом РС:

1) необходима минимизация вероятности "пропуска цели" (сегмент паузы определен как сегмент, содержащий полезную составляющую) для случая, когда на вход системы фильтрации в течение длительного интервала наблюдения поступает последовательность отсчетов РС, содержащих только акустическое воздействие;

2) необходима минимизация вероятности "ложных тревог" (сегмент, содержащий полезную составляющую, как сегмент паузы) для случая обработки длительного интервала наблюдения, содержащего как зашумленные сегменты пауз, так и зашумленные сегменты активных участков РС;

3) высокое быстродействие.

Условно провести следующую общую классификацию методов выделения паузы в обрабатываемом РС [89]:

- методы, основанные на статистических моделях воздействий;
- методы, основанные на отдельных характеристиках речи.

Однако современные детекторы активности речи опираются на комплексный подход и используют множество различных параметров [234]. Вследствие чего можно провести классификацию по используемым параметрам:

- 1) методы, использующие количество знакоперемен на сегменте локальной стационарности РС и пороговое значение мгновенной энергии сегмента;
- 2) методы на основе свойств автокорреляционной функции (АКФ);
- 3) методы, использующие особенности спектральных характеристик РС;
- 4) методы, основанные на исследовании поведения мощности на узкополосном отрезке частот – коэффициенты районной мощности;
- 5) методы, использующие кепстральный анализ РС;
- 6) методы на основе разности спектральных частот;
- 7) методы на основе оценки статистик высшего порядка.

Стоит отметить, что тот или иной метод обладает совокупностью собственных достоинств и недостатков [184]. Так, например, методы под пунктом 1 просты

в реализации, обладают низкой вычислительной сложностью, однако требуют априорных сведений об ОСШ для определения пороговых значений мгновенной энергии сегмента, ориентированы на аппроксимацию воздействия белым гауссовским шумом (БГШ), неэффективны при реальных акустических воздействиях (диапазон входных значений ОСШ от 40 до 15 дБ).

Методы на основе анализа АКФ также ориентированы на БГШ, но дополнительно учитывают корреляционные свойства вокализованных сегментов обрабатываемого РС, вследствие чего имеют значительные трудности в вопросах выделения невокализованных элементов речи (диапазон входных значений ОСШ для случая БГШ от 40 до 10 дБ). Методы под пунктами 3, 4, 5, 6 используют большинство показателей, тем или иным образом связанных со спектральными характеристиками РС, требуют большого количества априорных сведений о характере и уровне, как воздействий, так и РС в отсутствии зашумления, однако имеют широкий диапазон входных ОСШ от 40 до 0 дБ. Методы на основе оценок статистик высокого порядка имеют широкий диапазон входных значений ОСШ (от 40 до -5 дБ), требуют априорных сведений о законах распределения, как воздействия, так и активных речевых сегментов, обладают повышенной вычислительной сложностью, вследствие чего значительно возрастает алгоритмическая задержка. Исходя из вышеперечисленного, можно сделать вывод о наличии частной научной задачей диссертационного исследования, заключающейся в разработке процедуры выделения сегментов паузы при обработке РС в УИАВ.

При этом класс методов выделения паузы на основе оценки статистик высоких порядков в обрабатываемом РС в условиях интенсивного зашумления имеет наиболее широкий диапазон ОСШ, что позволяет сделать вывод о необходимости совершенствования данного подхода в вопросах адаптации посредством преодоления частичной априорной неопределенности путем анализа статистических и параметрических особенностей сегментов РС. Данный факт, несомненно, указывает на необходимость более детального изучения особенностей РС в УИАВ с исследованием свойств реальных акустических воздействий, который наглядно можно проиллюстрировать (рисунок 4.4) (исследовались стандартные детекторы активности речи, применяемые в МПАТ) [234].

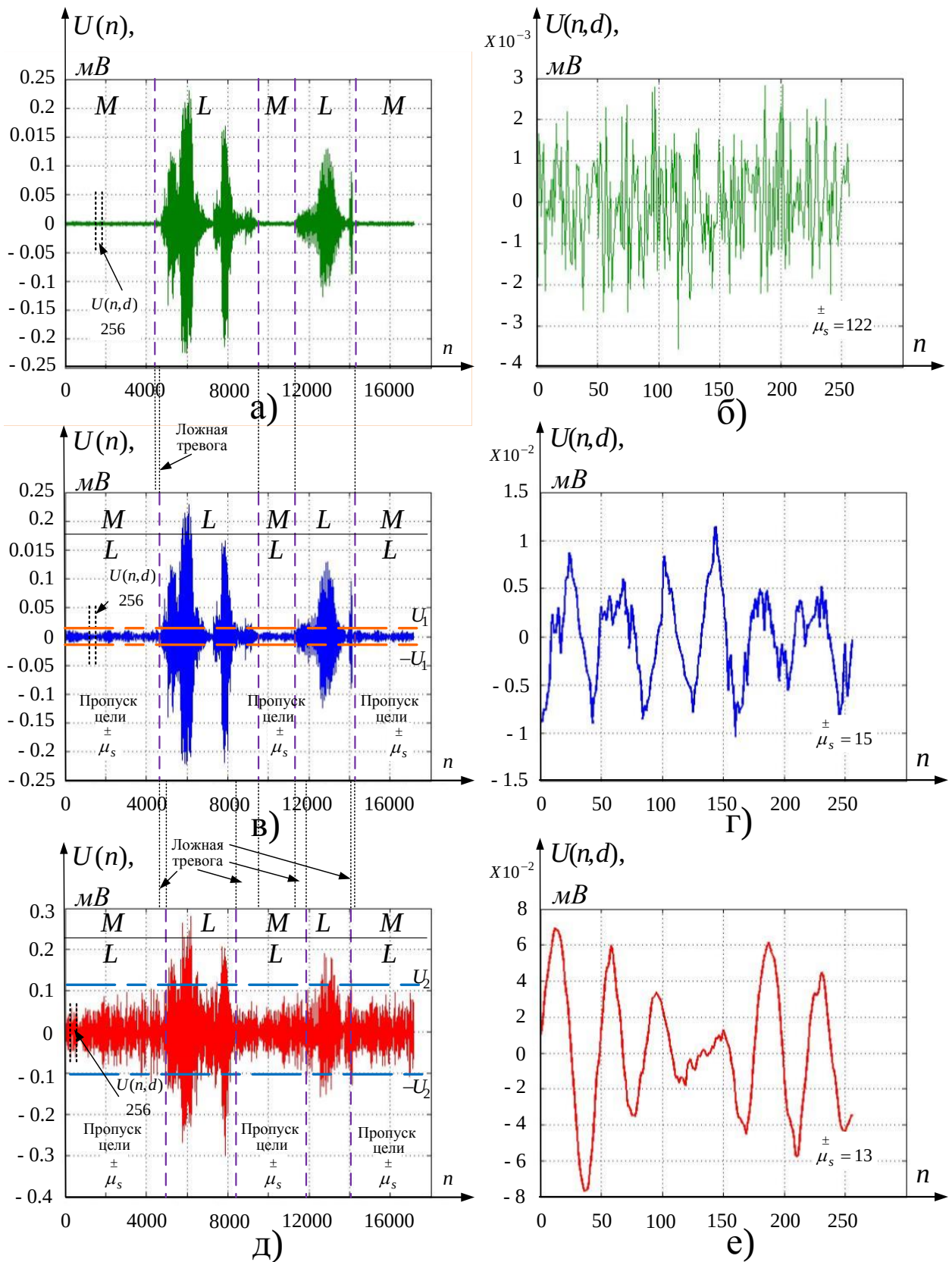


Рисунок 4.4 – Выделение сегментов паузы в существующих детекторах активности речи

Основными классификационными признаками сегментации служат интервальная оценка от $-U_1$ до U_1 (рисунок 4.4, в – РС в условиях акустического воздействия – шум проезжающего поезда ОСШ 20 дБ на интервале 2 секунды при частоте дискретизации 8000 Гц и 256 уровнях квантования, оценка ложных тревог через увеличение участков M) пороговой амплитуды или количество знакоперемен $\pm \mu_s$ на сегменте локальной стационарности:

$$\begin{aligned} |U(n, d)| \leq U_1 &\Rightarrow \rho(n, d) = 1, \\ |U(n, d)| > U_1 &\Rightarrow \rho(n, d) = 0, \end{aligned} \quad (4.19)$$

$$\rho_s(d) = \sum_{n=0}^{N-1} \rho(n, d), \quad (4.20)$$

$$\rho_s(d) \geq \rho_o. \quad (4.21)$$

В соответствии с условиями (4.19), отсчету $U(n, d)$ присваивается признак паузы $\rho(n, d) = 1$, в противном случае $\rho(n, d) = 0$ (т. е. данная реализация *Voice Activity Detector* (VAD) ориентирована на пороговое значение мгновенной энергии сегмента локальной стационарности). Суммарный коэффициент на сегменте d показывает общее количество отсчетов, удовлетворяющих пороговому условию (4.20). Вследствие этого очевидно появление в оценке порогового значения ρ_o , зависящего от количества отсчетов на сегменте РС, и при выполнении (4.21) сегмент определяется как пауза.

В [142] предложен новый метод и показаны решения по адаптации порогового значения U_2 , получаемого путем увеличения U_1 (4.19) для случая сложной помеховой обстановки. Значительное увеличение U_2 ведет к увеличению участков M , потенциально определяемых как пауза, и влечет за собой ошибку второго рода вида – "ложная тревога", когда активные составляющие обрабатываемого РС будут ошибочно иденти-

фицированы как сегменты пауз. Это приводит к ошибочной оценке параметров воздействия (рисунок 4.4, *д* – РС в условиях акустического воздействия – шум вертолета ОСШ 0 дБ на интервале 2 секунды при частоте дискретизации 8000 Гц и 256 уровнях квантования, оценка ложных тревог через увеличение участков M).

С учетом этого недостатка в [142] предложены решения поиска сегмента паузы во временной области РС путем подсчета количества перемен знака на сегменте $\mu_s^{\pm}(d)$ относительно амплитудной точки "0", основанного на предположении о шумовом воздействии, близком по своей природе к квазибелому гауссовскому шуму (КБГШ) (рисунок 4.4, *а* – РС, зашумленный КБГШ ОСШ 30 дБ; рис. 4.4, *б* – оценка количества перемен знака на сегменте паузы):

$$\begin{aligned} \mu_s^{\pm}(d) &\leq \mu_s^0 ; \\ \mu_s^0 &\cong N / 2 . \end{aligned} \quad (4.22)$$

Пороговое значение μ_s^0 зависит от количества отсчетов на сегменте локальной стационарности РС. При выполнении условия (4.22) сегмент определяется как пауза. Подход, основанный на использовании количества перемен знака, имеет высокую точность в диапазоне ОСШ до 20 дБ для случая, например, случайных флуктуаций мембраны микрофона МПАТ, когда искажение исходного сигнала, можно в приближенной степени аппроксимировать воздействием КБГШ. Однако для реальных акустических воздействий данный подход повышает вероятность ошибки второго рода в виде "пропуска цели", когда интенсивно зашумленная пауза РС будет принята за активный участок [54]. Это приводит к отсутствию оценки характеристик воздействия (рисунок 4.4, *в*, *г*, *д*, *е* – оценка количества перемен знака на сегменте локальной стационарности, вследствие чего все сегменты обрабатываемого РС определены как активные участки – L). Исходя из проведенной оценки детекторов активности речи, используемых в МПАТ, можно сделать вывод о том, что простота их реализации отрицательно скажется на функционировании всей системы фильтрации РС в УИАВ. Также в УИАВ выражение (4.12) уже не носит

характер приблизительного равенства, так как часть шума во взаимодействии с РС попадает в спектр мощности РС:

$$U(n, d) = S(n, d) + E(n, d) \Leftrightarrow U^2(n, d) = S^2(n, d) + E^2(n, d) + [2 \cdot E(n, d) \cdot S(n, d)] \Rightarrow \\ \Rightarrow A_{SS}(k, d) \neq \sqrt{|A_U^2(k, d) - \langle A_E^2(k, d) \rangle|}.$$

Решение данного несоответствия возможно согласно следующему выражению (прямое вычитание в спектрах амплитуд Фурье):

$$A_{SS}(k, d) \approx |A_U(k, d) - \langle A_E(k, d) \rangle|; \quad (4.23)$$

Выражение (4.23) позволяет осуществлять фильтрацию в задаче разделения сложных сигналов, однако неточная оценка спектра амплитуд воздействия ведет к значительным ошибкам при вычитании спектров. Вследствие чего выполнение спектрального вычитания требует дополнительной коррекции и более точной оценки спектральных характеристик акустического воздействия.

4.3 Математическая модель системы фильтрации речевого сигнала на основе применения полиспектрального анализа

На основании вышеперечисленных недостатков существующей модели системы фильтрации с использованием спектрального вычитания можно сделать вывод о необходимости ее совершенствования с целью повышения качества РС в условиях интенсивных акустических воздействий (УИАВ – ОСШ менее 15 дБ), что может быть достигнуто с использованием следующего подхода – это использование модели системы фильтрации РС на основе полиспектрального анализа [11, 23, 50, 54], согласно которой РС представляется аддитивной смесью речи и побочных акустических воздействий, тогда сегмент зашумленного РС может быть представлен суммой компонент, носящих как гауссов $K_g(t)$, так и негауссов характер $K_{ag}(t)$:

$$U(t, d) = S(t, d) + E(t, d) , \quad (4.24)$$

$$U(t, d) = \sum_{g=1}^{g=G} K_g(t, d) + \sum_{ag=1}^{ag=aG} K_{ag}(t, d) , \quad (4.25)$$

где $U(t, d)$ – обрабатываемый сегмент непрерывного РС, $S(t, d)$ – речевая волна, $E(t, d)$ – аналоговый акустический шум.

Закон распределения мгновенных значений РС на сегменте локальной стационарности, представленного выражениями (4.24, 4.25), в условиях, когда энергетически шум равен или превышает РС, стремится, в первом приближении, к закону распределения акустического воздействия (при учете симметрии закона распределения мгновенных значений акустического воздействия), однако имеет отличия от него (рисунок 4.5).

Описание реального случайного процесса может быть выполнено следующим образом [115]:

$$W_{\zeta}(u) = \frac{1}{2\pi} \cdot \int_{-\infty}^{\infty} \Theta_{\xi}(v) \exp[jvu] dv , \quad (4.26)$$

$$\Theta_{\xi}(v) = \int_{-\infty}^{\infty} W_{\zeta}(u) \exp[-jvu] du , \quad (4.27)$$

$W_{\zeta}(u)$ – ФПВ случайной величины; $\Theta_{\xi}(v)$ – характеристическая функция.

Данную пару еще называют "статистической парой преобразования Фурье", что позволяет говорить о характеристической функции как о тождественном представлении ФПВ [115].

Так как характеристическая функция тождественно представляет ФПВ, то все параметры функции $W_{\zeta}(u)$ можно получить из $\Theta_{\xi}(v)$, например, $a_x[\zeta]$ – начальные моменты случайной величины ζ :

$$a_x[\zeta] = j^{-n} \left(\frac{\partial^n \Theta_{\xi}(v)}{\partial v^n} \right)_{v=0} \equiv j^{-x} \cdot \Theta_{\xi}^x(0) . \quad (4.28)$$

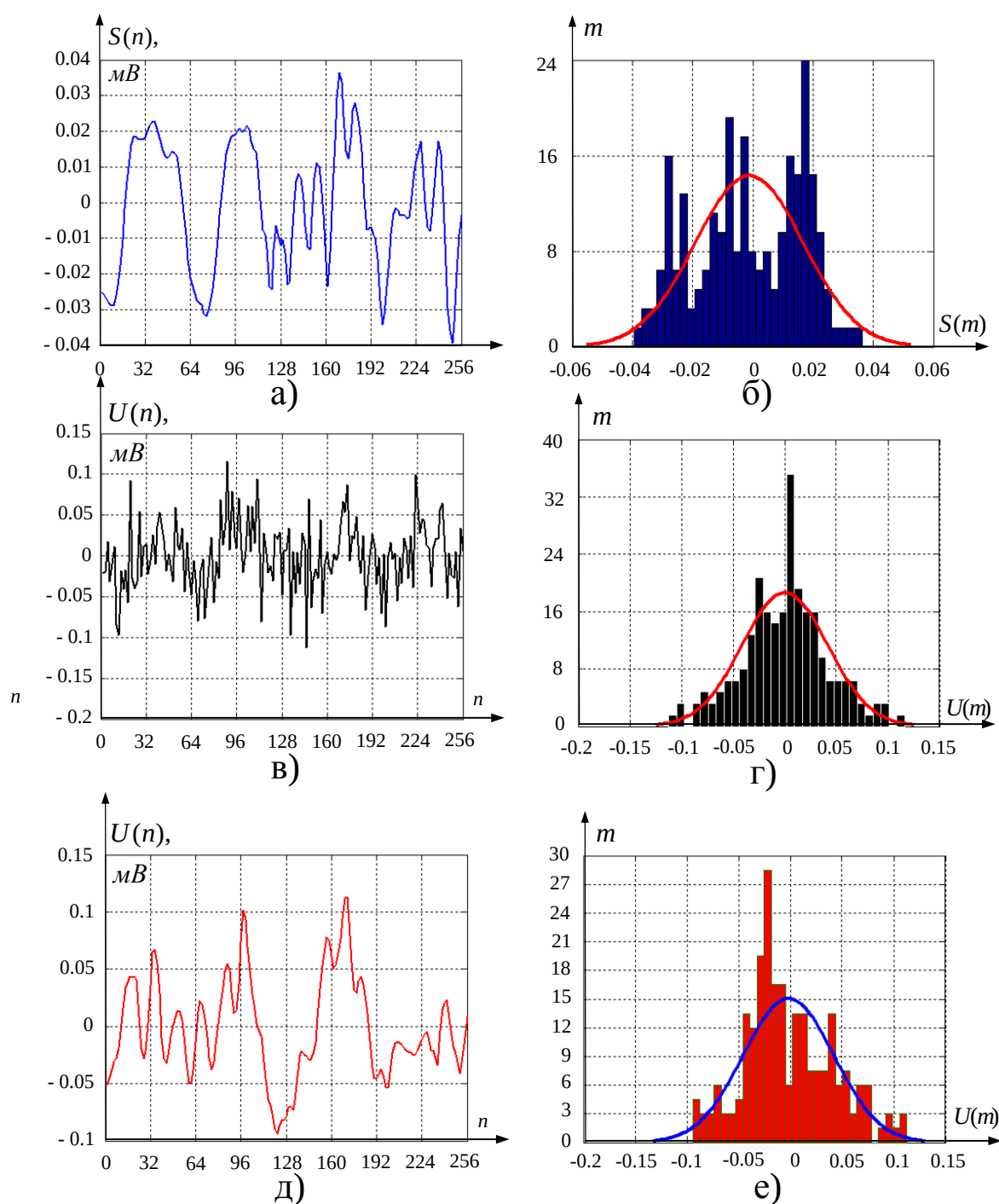


Рисунок 4.5 – Свойства асимметрии закона распределения

мгновенных значений РС на активных участках:

- а), б) сегмент РС при отсутствии зашумления – высокая асимметричность;
 в), г) сегмент, зашумленный КБГШ ОСШ 5 дБ – присутствие асимметричности со значительным ослаблением;
 д), е) сегмент, зашумленный шумом двигателя ОСШ 5 дБ – присутствие асимметрии с незначительным ослаблением.

В обратном случае ХФ можно записать с помощью степенного ряда, членами которого являются моменты случайной величины:

$$\Theta_{\xi}(v) = 1 + \sum_{x=1}^{\infty} \frac{a_x[\xi]}{x!} \cdot (jv)^x. \quad (4.29)$$

Если же записать ХФ как некую экспоненциальную функцию от определенной функции $B(v)$:

$$\Theta_{\xi}(v) = \exp[B(v)], \quad (4.30)$$

тогда в случае нахождения обратной функции для (4.30) можно записать:

$$B(v) = \ln \Theta_{\xi}(v) = \sum_{y=1}^{\infty} \frac{\chi_y}{y!} \cdot (jv)^y, \quad (4.31)$$

вследствие чего саму ХФ можно представить через члены ряда функции $B(v)$:

$$\Theta_{\xi}(v) = \exp\left[\sum_{y=1}^{\infty} \frac{\chi_y}{y!} \cdot (jv)^y\right], \quad (4.32)$$

где χ_y – коэффициенты ряда функции $B(v)$ и носят название кумулянтов или семинвариантов. Набор кумулянтов описывает статистические характеристики случайной величины и в совокупности служит тождественным представлением ее закона распределения.

Анализируя (4.32) и (4.29), можно сделать вывод, что кумулянты и моменты случайной величины связаны между собой [76, 77, 115]:

$$\begin{aligned} \chi_1 &= a_1 \\ \chi_2 &= a_2 - a_1^2 \\ \chi_3 &= a_3 - 3a_1 a_2 + a_1^3 \\ \chi_4 &= a_4 - 3a_2^2 - 4a_1 a_3 + 12a_1^2 a_2 - 6a_1^4 \\ &\dots \end{aligned} \quad (4.33)$$

Аналогично показана и обратная связь:

$$\begin{aligned}
a_1 &= \chi_1 \\
a_2 &= \chi_2 + \chi_1^2 \\
a_3 &= \chi_3 + 3\chi_1 \chi_2 + \chi_1^3 \\
a_4 &= \chi_4 + 3\chi_2^2 + 4\chi_1 \chi_3 + 6\chi_1^2 \chi_2 + \chi_1^4 \\
&\dots
\end{aligned} \tag{4.34}$$

Необходимо отметить, что коэффициенты χ_3 и χ_4 в выражении (4.33) характеризуют асимметрию и эксцесс ФПВ, кроме того, для *гауссова процесса* данные коэффициенты *равны нулю*.

Непосредственно для практического применения, раскрываемого в исследовании, наиболее важна ФПВ негауссовского процесса $W_{aG_\zeta}(u)$, получаемая путем обратного преобразования Фурье от (4.32):

$$W_{aG_\zeta}(u) = W_{G_\zeta}(u) + \sum_{y=3}^{\infty} (-1)^y \cdot \frac{\chi_y}{y!} \cdot W_{G_\zeta}^{(y)}(u) , \tag{4.35}$$

где $W_{G_\zeta}(u)$ – ФПВ, носящая гауссовский характер; $\sum_{y=3}^{\infty} (-1)^y \cdot \frac{\chi_y}{y!} \cdot W_{G_\zeta}^{(y)}(u)$ – сумма произведений кумулянтных функций на производные y -го порядка $W_{G_\zeta}(u)$.

Полнота представления и описания стационарного сигнала конечной мощности может быть достигнута не только совокупностью центральных моментов, кумулянтов, но и *набором полиспектров, которые являются Фурье-отображением кумулянтных функций*:

$$\begin{aligned}
A_U^2(k) &\leftrightarrow \chi_2, \\
\dot{B}_U^3(k) &\leftrightarrow \chi_3, \\
&\dots, \\
\dot{C}_U^y(w_1, \dots, w_{n-1}) &\leftrightarrow \chi_y, \\
&\Downarrow \\
W_{aG_\zeta}(u) &\equiv \left\{ A_U^2(k), \dot{B}_U^3(p, q), \dots, \dot{C}_U^y(w_1, \dots, w_{n-1}) \right\},
\end{aligned} \tag{4.36}$$

где $A_U^2(k)$ – спектр мощности; $\dot{B}_U^3(p, q)$ – биспектр; $\dot{C}_U^y(w_1, \dots, w_{n-1})$ – мультиспектр y -го порядка.

Исходя из свойств асимметрии значений активных участков РС в условиях зашумления можно сделать вывод, что для анализа РС целесообразно использовать совокупную оценку спектра мощности и биспектра, которая несет в себе более полную информацию о характере ФПВ обрабатываемого сигнала [124]. Биспектр дискретного сигнала – есть двумерное Фурье-отображение тройной автокорреляционной функции (ТАКФ) $R_U(a, b)$, которая раскрывает физическое значение бикорреляции и описывающей в статистическом смысле сечение случайного процесса (СП) по кумулянту 3-го порядка – коэффициенту асимметрии χ_3 (4.33, 4.36) (для эргодического СП) при $n = 0, 1, 2, \dots, N - 1$:

$$R_U(a, b) = \sum_{n=0}^{N-1} U(n) \cdot U(n+a) \cdot U(n+b), \tag{4.37}$$

$$R_U(a) = \sum_{n=0}^{N-1} U(n) \cdot U(n+a); \tag{4.38}$$

$$R_U(a) = R_U(-a). \tag{4.39}$$

Как и классическая функция автокорреляции (4.38) отличается свойством вида (4.39), ТАКФ также обладает следующим важным для практики свойством симметрии:

$$R_U(a, b) = R_U(b, a) = R_U(b - a, -a) = R_U(a - b, -b) = R_U(-a, b - a) = \dots \quad (4.40)$$

Из определений, сформулированных в [174] биспектр определяется следующим образом:

$$\dot{B}_U^3(p, q) = \sum_{a=-N+1}^{N-1} \sum_{b=-N+1}^{N-1} R_U(a, b) \cdot \exp[-j2\pi(ap + bq)]. \quad (4.41)$$

При этом биспектр – это комплексная функция двух переменных p и q :

$$\dot{B}_U^3(p, q) = B_U^3(p, q) \cdot \exp[j\gamma_U(p, q)], \quad (4.42)$$

где $B_U^3(p, q)$ – биамплитуда, $\gamma_U(p, q)$ – бифаза.

С учетом симметрии ТАКФ и биспектра (4.37, 4.41) преобразуем к виду:

$$\dot{B}_U^3(p, q) = \sum_{a=0}^{N-1} \sum_{b=0}^{N-1} R_U(a, b) \cdot \exp[-j2\pi(ap + bq)]. \quad (4.43)$$

В работе [70] приводятся основные свойства биспектра, такие как:

1) биспектр – периодическая функция с периодом, равным 2π :

$$\dot{B}_U^3(p, q) = \dot{B}_U^3(p + 2\pi, q + 2\pi), \quad (4.44)$$

2) биспектр – это симметричная функция на биспектральной плоскости (плоскости двух независимых частотных переменных p и q):

$$\begin{aligned} \dot{B}_U^3(p, q) &= \dot{B}_U^3(q, p) = \dot{B}_U^{3*}(-q, -p) = \dot{B}_U^{3*}(-p, -q) = \dot{B}_U^{3*}(-p - q, q) = \\ &= \dot{B}_U^3(p, -p - q) = \dot{B}_U^3(-p - q, p) = \dot{B}_U^3(q, -p - q) = \dots \end{aligned} \quad (4.45)$$

3) Биспектр обладает свойством подобия, что явно раскрывает прямую взаимосвязь между биамплитудой и спектром амплитуд Фурье, а также взаимосвязь бифазы и спектром фаз Фурье:

$$B_U^3(p, q) = A_U(p) \cdot A_U(q) \cdot A_U(p + q), \quad (4.46)$$

$$\gamma_U(p, q) = F_U(p) + F_U(q) - F_U(p + q). \quad (4.47)$$

Свойства, показанные в выражениях (4.46, 4.47), применяются в итерационных рекурсивных алгоритмах восстановления сигнала по оценке биспектра, а также непосредственно при прямом методе получения значений биспектра. Оценка биспектра есть совокупность двумерного преобразования Фурье от ТАКФ, полученной в результате усреднения по ансамблю временных реализаций СП [16]. Однако рассмотрение РС как квазиэргодического процесса на участках локальной стационарности дает возможность перейти к вычислению биспектра по конкретной реализации РС.

$$\dot{B}_U^3(p, q, d) \approx \dot{B}_S^3(p, q, d) + \dot{B}_E^3(p, q, d), \quad (4.48)$$

где $\dot{B}_U^3(p, q, d)$ – оценка биспектра РС на сегменте d ;

$\dot{B}_S^3(p, q, d)$ – оценка биспектра РС в отсутствии зашумления на сегменте d ;

$\dot{B}_E^3(p, q, d)$ – оценка биспектра акустического воздействия на сегменте d .

В условиях «стандартного» зашумления, т.е. в случае акустического воздействия, представляющей собой процесс вида белого шума с нулевым средним значением, биспектр искаженного РС не дает дополнительной информации о характере ФПВ воздействия, так как *статистики высоких порядков гауссовских процессов стремятся к нулю*, однако в этом случае увеличивается объем данных, анализируемых на сегменте, а также позволяет в некоторой степени ослабить шумоподобные компоненты при использовании методов накопления и усреднения. Если же сигнал акустического воздействия отличается либо по уровню среднего значения, либо по своему распределению от нормального закона, то совокупная оценка мультиспектров укажет на факт наличия сегмента обработки, носящего негауссовский характер. При этом оценка биспектральной плотности РС позволяет оценить

асимметрию ФВП, с дальнейшим использованием полученной информации для оценки характеристик акустического воздействия, так как совокупная оценка полиспектров более точно описывает статистические свойства РС в УИАВ.

В случае акустического воздействия высокого уровня, представляющего собой процесс негауссовского вида (4.24, 4.25), использование известной математической модели с применением спектрального вычитания, которая, в свою очередь, ориентирована на "нормальный" закон распределения, дает значительные ошибки при оценке усредненных спектральных характеристик [39].

В условиях акустического воздействия, представляющего собой процесс с гауссовским характером закона распределения мгновенных значений, биспектр зашумленного РС показывает, в большей степени, информации о характере ФПВ речевого сигнала воздействия, так как *статистики высоких порядков гауссовских процессов стремятся к нулю*. Тогда при обработке сегмента, содержащего полезную речевую составляющую, закон которой отличается от нормального, совокупная оценка мультиспектров укажет на факт наличия сегмента обработки, носящего негауссовский характер.

Исследование корреляции высоких порядков обрабатываемых сегментов РС в условиях акустического зашумления позволяет провести правильную сегментацию с целью выделения паузы в РС [56]. При этом анализ энергетических характеристик во временной области сведен к минимуму [142]. На основании вышеизложенного в рамках диссертационного исследования предложена модель системы фильтрации на основе полиспектрального анализа, которая представлена на (рисунке 4.6).

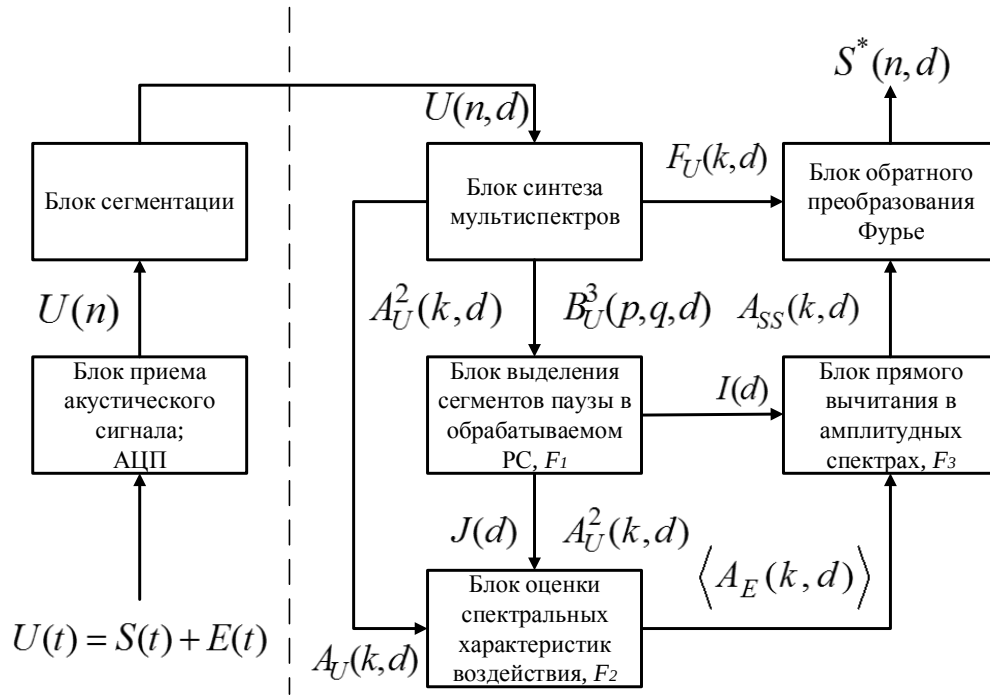


Рисунок 4.6 – Функциональная модель системы фильтрации речевого сигнала на основе полиспектрального анализа

Представим данную модель в виде следующего математического выражения:

$$U(n, d) \xrightarrow{\Pi\{F_1, F_2, F_3\}} S^*(n, d), \quad (4.49)$$

где Π – оператор системы фильтрации РС; F_1 – функционал выделения сегментов паузы; F_2 – функционал оценки спектральных характеристик акустического воздействия; F_3 – функционал шумоподавления на основе прямого вычитания в амплитудных спектрах.

При этом РС на выходе системы фильтрации $S^*(n, d)$ определяется как:

$$S^*(n, d) = \frac{1}{N} \cdot \sum_{k=0}^{N-1} \dot{A}_{SS}(k, d) \cdot e^{kn}, \quad (4.50)$$

где $\dot{A}_{SS}(k, d)$ – комплексный спектр Фурье сегмента на выходе системы фильтрации:

$$\dot{A}_{SS}(k, d) = A_{SS}(k, d) \cdot \exp[F_U(k, d)], \quad (4.51)$$

$$A_{SS}(k, d) = F_3\{A_U(k, d), \langle A_E(k, d) \rangle, I(d)\}, \quad (4.52)$$

где $F_U(k, d)$ – оценка спектра фаз зашумленного сегмента обрабатываемого РС;
 $A_U(k, d)$ – оценка спектра амплитуд зашумленного сегмента d обрабатываемого РС;
 $\langle A_E(k, d) \rangle$ – усредненная оценка спектра амплитуд воздействия для сегмента d , определяемая при $d = 1, 2, \dots, D$, $r = 0, 1, \dots, M$ по следующему выражению:

$$\langle A_E(k, d) \rangle = F_2 \{ A_U(k, r), A_U^2(k, r), J(d) \}, \quad (4.53)$$

где $A_U^2(k, d)$ – совокупность оценок спектра мощности для D – выборки обрабатываемых сегментов, M из которых сегменты паузы РС и L – сегментов активных участков, содержащих полезную составляющую; $A_U(k, r)$ – совокупность оценок спектра амплитуд M – выборки сегментов паузы обрабатываемого РС; $J(d)$ – бинарная функция признака паузы, определяемая как:

$$\begin{aligned} J(d) &= F_1 \{ A_U^2(k, d), B_U^3(p, q, d) \}, \\ d \in M &\Rightarrow J(d) = 1, \\ d \in L &\Rightarrow J(d) = 0, \end{aligned} \quad (4.54)$$

где $B_U^3(p, q, d)$ – совокупность оценок биамплитуды для D – выборки обрабатываемых сегментов, получаемой согласно прямому методу:

$$B_U^3(p, q, d) = A_U(p, d) \cdot A_U(q, d) \cdot A_U(p + q, d), \quad (4.55)$$

тогда коэффициент усреднения – $J(d)$ (4.54) определяется согласно следующему правилу [54, 56, 142]:

$$\begin{aligned} J(d) = 1 &\Rightarrow I(d) = 1, \\ J(d) = 0 &\Rightarrow I(d) = \frac{1}{2}. \end{aligned} \quad (4.56)$$

Стоит отметить и высокую вариативность РС в УИАВ, вследствие чего даже при достаточно точной усредненной оценке спектральных характеристик акустического воздействия провести шумоподавление на основе спектрального вычитания без возникновения искажений не представляется возможным. Таким образом, при обработке

РС активно используются процедура интерполяционного сглаживания и низкочастотной фильтрации. Исследования показали, что для устранения данного недостатка на выходе цифровых систем фильтрации с использованием спектрального вычитания достаточно применения процедуры "усреднения паузы" [56, 142], что позволяет дополнительно ослабить шум на 5-6 дБ на сегментах, не содержащих полезной составляющей, используя при этом остаточный шум для маскировки искажений и передачи окружающей обстановки.

Учет психоакустического восприятия РС при реализации комплексного подхода к вопросу шумоподавления в УИАВ позволяет более эффективно проводить процедуру фильтрации [7, 107, 220]. При этом недостаток существующих алгоритмов спектрального вычитания, заключающийся в нелинейных искажениях, частично устраняется.

В соответствии с вышесказанным целевая функция требует уточнения задачи выявления взаимосвязи между усредненной оценкой спектральных характеристик акустического воздействия и сигнала на выходе системы фильтрации. Тогда математическую интерпретацию линейной разницы ОСШ ΔG можно представить следующими выражениями:

$$\Delta G = G^{out} - G^{in}, \quad (4.57)$$

$$G^{out} = \frac{1}{D} \cdot \sum_{d=1}^{d=D} 10 \cdot \lg \left(\sum_{n=0}^{n=N-1} \frac{[S(n,d)]^2}{[S^*(n,d) - S(n,d)]^2} \right), \quad (4.58)$$

$$G^{in} = \frac{1}{D} \cdot \sum_{d=1}^{d=D} 10 \cdot \lg \left(\sum_{n=0}^{n=N-1} \frac{[S(n,d)]^2}{[U(n,d) - S(n,d)]^2} \right), \quad (4.59)$$

где G^{out} – ОСШ на выходе системы фильтрации; G^{in} – ОСШ на входе системы фильтрации; D – количество сегментов локальной стационарности в РС.

Анализ (4.57, 4.58) показывает максимум, который достигим при разности $S^*(n,d) - S(n,d) = 0$ для всех обрабатываемых сегментов. Кроме того, анализ (4.49–4.56) выявил поэтапность решения задачи шумоподавления. Таким образом,

выполнение функционала F_3 зависит от F_2 , который в свою очередь зависит от F_1 . Вследствие этого главной задачей идеализированной системы фильтрации является сведение усредненной оценки $\langle A_E(k, d) \rangle$ к истинной оценке спектра амплитуд Фурье акустического воздействия $A_E(k, d)$ для всех k, d при выполнении условия:

$$A_E(k, d) - \langle A_E(k, d) \rangle = 0. \quad (4.60)$$

При выполнении (4.60) необходимо отметить:

1) получение точной оценки спектральных характеристик воздействия для каждого обрабатываемого сегмента зашумленного РС ведет к достижению *максимума в рамках предложенной модели*;

2) вырождение процедуры усреднения сегментов паузы с целью формирования остаточного шума ведет к возникновению воспринимаемых слуховым аппаратом абонента незначительных искажений (для случая УИАВ).

При воздействии реальных шумов выполнение условия (4.60) труднодостижимо, что и приводит к выполнению полного перечня процедур в модели. С другой стороны, анализ (4.57-4.59) требует дополнительного введения понятия шума на выходе системы фильтрации РС, определяемого как:

$$E^*(n, d) = S^*(n, d) - S(n, d), \quad (4.61)$$

где $E^*(n, d)$ – шум на выходе системы фильтрации для сегмента d РС.

В свою очередь, $E^*(n, d)$ можно представить следующим образом:

$$E^*(n, d) = \begin{cases} E_{rez}^*(n, d) + E_{mat}^*(n, d), & d \in L; \\ \frac{1}{2} \cdot (E_{rez}^*(n, d) + E_{mat}^*(n, d)), & d \in M, \end{cases} \quad (4.62)$$

где $E_{rez}^*(n, d)$ – остаточный шум на выходе системы; $E_{mat}^*(n, d)$ – собственные шумы системы (шумы округлений, шум квантования и др.).

Анализ (4.61, 4.62) выявляет дополнительные ограничения в модели:

$$W - \xi \leq E_{mat}^*(n, d) \leq W + \xi; \quad (4.63)$$

$$E_{rez}^*(n, d) \leq E_{rez}^o(n, d), \quad (4.64)$$

где W – математическое ожидание вычислительной стабильности цифровой системы; ξ – параметр доверительного интервала вычислительной стабильности цифровой системы; $E_{rez}^o(n, d)$ – пороговое значение остаточного шума на выходе системы, вносимого при проведении фильтрации РС.

Тогда при учете дополнительных ограничений, математическую модель системы фильтрации можно представить в виде:

$$\begin{aligned} U(n, d) &\xrightarrow{\Pi\{F_1, F_2, F_3\}} S^*(n, d), \\ E^*(n) &\rightarrow \min \Leftrightarrow \Delta G \rightarrow \max, \\ \left. \begin{aligned} E_{rez}^*(n, d) &\leq E_{rez}^o(n, d) \\ W - \xi &\leq E_{mat}^*(n, d) \leq W + \xi \end{aligned} \right\} &\Leftrightarrow \frac{Q^{out} \geq Q^o \quad \forall Q^{in} \leq Q^o}{Q^{out} \geq Q^{in} \quad \forall Q^{in} > Q^o}. \end{aligned} \quad (4.65)$$

Стоит отметить, что исследование стабильности системы фильтрации с использованием спектрального вычитания проводились многими исследователями [184], вследствие чего оценка данного параметра для $E_{rez}^*(n, d)$ не требуется. Однако остается вопрос относительно стабильности собственных шумов предложенной системы фильтрации $E_{mat}^*(n, d)$, которые при дополнительном введении оценки биспектральной плотности может иметь сингулярный характер по причине наличия нелинейных связей при оценке мультиспектров.

Для решения этой задачи при моделировании системы фильтрации РС на основе полиспектрального анализа рассматривались вопросы чувствительности и адекватности модели, проверка модели на акустическое воздействие квазистационарного характера, а также на вносимую погрешность при проведении вычислений, которая определяет шум $E_{mat}^*(n, d)$, связанный с возрастанием вычислительной сложности и размерности пространства, а также других ограничений, определяемых особенностями ЦОС.

Для проверки адекватности модели при воздействии различных РС используются тестовые фразы согласно [84]. Моделирование выполнено в программной среде MATLAB [26]. Особенности реализации представлены в Приложении 1.

При проверке погрешности при вычислениях используется дисперсия $\chi_2(0, z)$ и среднеквадратическое отклонение (СКО) $\sigma(z)$ определяемые согласно следующим выражениям:

$$\chi_2(0, z) = \frac{1}{D^*(N-1)-1} \cdot \sum_{n=0}^{n=D^*(N-1)} [U^*(n, z) - U(n, z)]^2, \quad (4.66)$$

$$\sigma(z) = \sqrt[2]{\chi_2(0, z)}, \quad (4.67)$$

где $U(n, z)$ – последовательность отсчетов зашумленной тестовой фразы случайными уровнем зашумления и видом воздействия для z -го опыта, $U^*(n, z)$ – реакция системы на входное воздействие $U(n, z)$ в виде последовательности отсчетов на выходе системы, определяющей все этапы обработки, связанные с введением в систему фильтрации оценки биспектральной плотности (получение оценки биспектра, восстановление оценок спектров Фурье по оценке биспектра, получение сигнала во временной области путем обратного преобразования Фурье).

Тогда оценка математического ожидания вычислительной стабильности W (4.63) является оценкой математического ожидания среднеквадратического отклонения (4.66) по достаточно большому количеству опытов:

$$W = \frac{1}{Z} \cdot \sum_{z=1}^{n=Z} \sigma(z), \quad (4.68)$$

Экспериментальная проверка представленной математической модели проводилась на основе анализа влияния акустических воздействий различной природы (рисунки 4.7–4.9). РС, получаемый на выходе акустико-электрического преобразователя подвергался АЦП с частотой дискретизации 8 кГц, использовалось 256 уровней квантования. Приведенный к таким же характеристикам нормализованный акустический шум аддитивно смешивался с исходным РС для получения зашумленных реализаций РС.

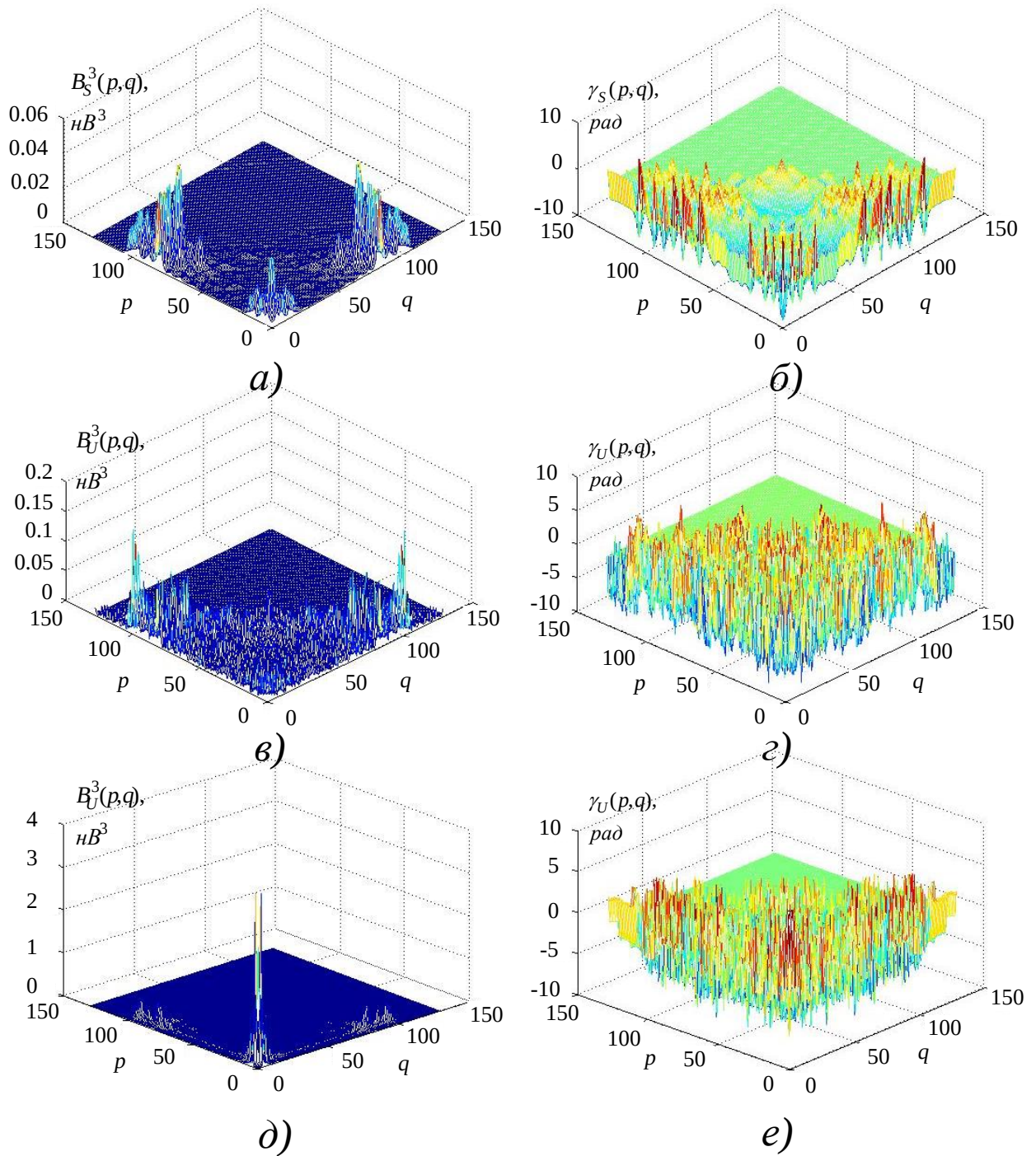


Рисунок 4.7 – Анализ зашумленного РС в биспектральной области:

- а) оценка биамплитуды сегмента РС в отсутствии зашумления;
- б) оценка бифазы сегмента РС;
- в) оценка биамплитуды зашумленного сегмента (КБГШ ОСШ 5 дБ);
- г) оценка бифазы зашумленного сегмента КБГШ ОСШ 5 дБ;
- д) оценка биамплитуды зашумленного сегмента (шум вертолета ОСШ 0 дБ);
- е) оценка бифазы зашумленного сегмента (шум вертолета ОСШ 0 дБ).

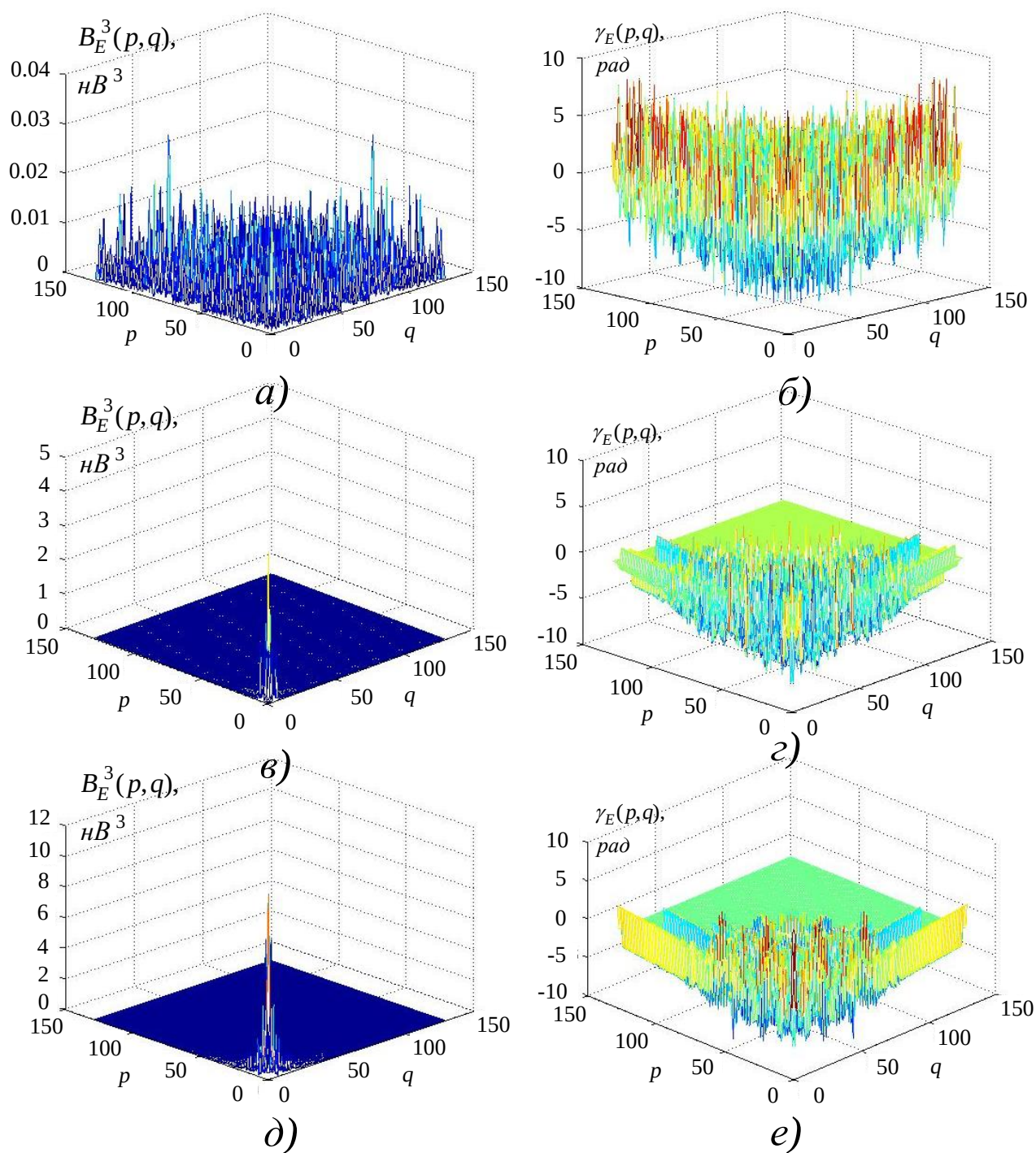


Рисунок 4.8 – Анализ акустических воздействий в биспектральной области:

- а) оценка биамплитуды сегмента паузы РС (КБГШ ОСШ 5 дБ);
- б) оценка бифазы сегмента паузы РС (КБГШ ОСШ 5 дБ);
- в) оценка биамплитуды сегмента паузы РС (шум вертолета ОСШ 0 дБ);
- г) оценка бифазы сегмента паузы РС (шум вертолета ОСШ 0 дБ);
- д) оценка биамплитуды сегмента паузы РС (шум ветра ОСШ 0 дБ);
- е) оценка бифазы сегмента паузы РС (шум ветра ОСШ 0 дБ).

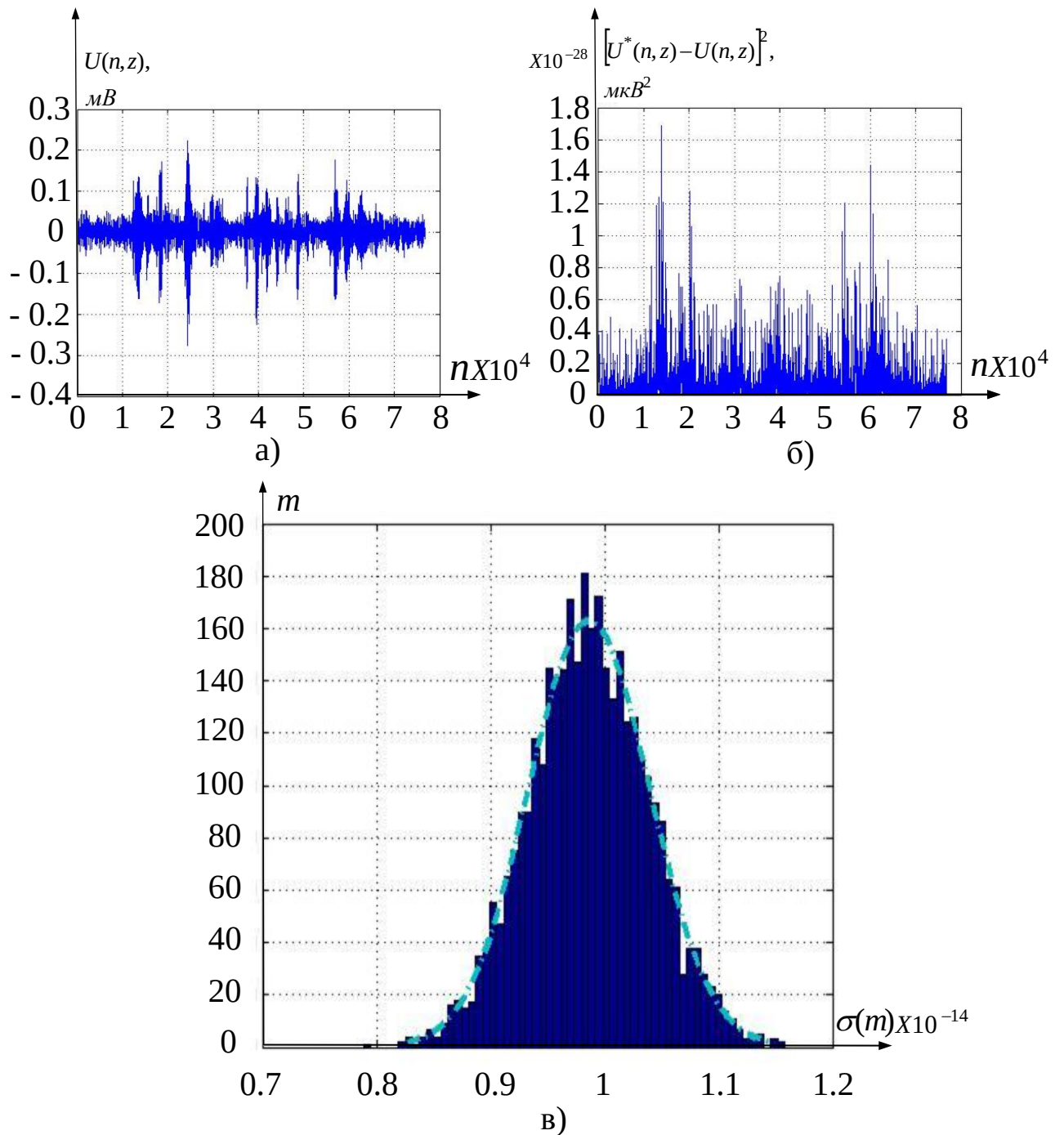


Рисунок 4.9 – Оценка вычислительной стабильности предложенной системы фильтрации РС на основе полиспектрального анализа:

- а) последовательность отсчетов, подаваемая на вход системы;
- б) оценка квадрата отклонения сигнала на выходе от сигнала на входе системы по отсчетам;
- в) гистограмма значений СКО по 3500 опытам.

Анализ (рисунка 4.7) позволяет сделать вывод о том, что дополнительное введение в предложенную модель системы фильтрации оценки биспектральной плотности не вносит существенных шумов вычислений и определяется ошибкой округления. Вследствие этого можно пренебречь данной составляющей шума на выходе системы фильтрации и уточнить выражение (4.49):

$$\begin{aligned}
 & U(n, d) \xrightarrow{\Pi\{F_1, F_2, F_3\}} S^*(n, d), \\
 & A_E(k, d) - \langle A_E(k, d) \rangle \rightarrow \min \Leftrightarrow E_{rez}^*(n, d) \rightarrow \min \Leftrightarrow \Delta G \rightarrow \max, \\
 & \frac{Q^{out} \geq Q^o \quad \forall Q^{in} \leq Q^o}{Q^{out} \geq Q^{in} \quad \forall Q^{in} > Q^o}, \tau \leq \tau^o.
 \end{aligned} \tag{4.69}$$

Анализ (рисунков 4.7–4.9) показывает, что КБГШ высокой интенсивности (рисунки 4.7, в, г) ведет к изменению исходной биспектральной плотности РС (рисунки 4.7, а, б), как в высокоплотностной, так и в низкоплотностной области, что влечет значительное снижение субъективной оценки качества РС. Кроме того, анализ реальных акустических воздействий показал, что разрушение исходной структуры РС менее значительны, чем для случая КБГШ, однако носят «индивидуальный» характер в зависимости от вида шума.

Также анализ низкоплотностной области биамплитуды РС (для случая реального акустического узкополосного воздействия в УИАВ (рисунки 4.7, д, е; рисунок 4.8, в – е) определяет перспективность оценки разных областей биамплитуды для эффективного разделения обрабатываемых сегментов РС на сегменты паузы и активные участки [16, 56, 142]. Квазистационарные акустические воздействия сложного характера и различной природы вносят изменения согласно своим статистическим и параметрическим особенностям. Это указывает на необходимость "индивидуальной" оценки каждого вида акустического воздействия или использования обобщенного подхода, учитывающего особенности каждого сигнала, что и позволяет сделать полиспектральный анализ.

Предложенная математическая модель системы фильтрации РС на основе полиспектрального анализа дополнительно учитывает оценку биспектра, а также

использует полученную информацию для оценки характеристик акустического воздействия и проведения шумоподавления на основе подхода, заключающегося в разделении РС с использованием вычитания в спектрах амплитуд. Предложенное техническое решение дает возможность разработать новый перспективный подход к реализации систем шумоподавления в УИАВ [141].

Информация, полученная из дополнительной оценки биспектра, носит качественный характер. Пояснить данный факт можно по выражению (4.29), носящему также название закона Эджворта [70, 115], анализ которого показывает, что по сравнению с существующими методами аппроксимации закона распределения РС при его зашумлении в совокупной оценке мультиспектров формируется на $\sum_{k=3}^{\infty} (-1)^k \frac{\chi_k}{k!} W_{G_\zeta}^{(k)}(u)$ больше качественной информации. Однако качество этой информации убывает с обратной факториальной зависимостью к количеству.

Также следует учитывать возрастание вычислительной сложности, которая зависит от количества отсчетов на сегменте $N - 1$ и носит степенной характер. Недостатки существующей модели системы фильтрации на основе спектрального вычитания во многом устранены, что позволяет разработать более эффективные алгоритмы обработки РС в УИАВ.

4.4 Выделение сегментов паузы и оценка спектральных характеристик помех

Анализ разработанной модели системы фильтрации в условиях зашумления с применением пошагового выполнения вложенных оптимизаций показывает, что для более точного оценивания спектральных характеристик акустического воздействия необходимо решить задачу выделения паузы в обрабатываемом РС. С учетом данного факта необходимо определить требования к разрабатываемому алгоритму выделения сегментов паузы.

Необходимо минимизировать вероятность «пропуска цели» P_{prc} (сегмент паузы определен как, содержащий полезную составляющую) для случая, когда на вход

системы обработки РС в течение длительного интервала наблюдения поступает последовательность отсчетов, содержащих только акустическое воздействие:

$$P_{prc} \rightarrow \min, \quad (4.70)$$

Широкий диапазон входных значений ОСШ от 30 до -5 дБ.

Процедура выделения паузы, должна учитывать возможность применения в условиях зашумления РС аддитивным КБГШ, т.е. должна выполнять разные операции для широкополосных и узкополосных акустических воздействий.

При этом необходимо также зафиксировать наименьшую вероятность «ложных тревог» P_{lt} (сегмент, содержащий полезную составляющую, определен как сегмент паузы) для случая обработки длительного интервала наблюдения, содержащего как зашумленные сегменты пауз, так и зашумленные сегменты активных участков РС (снижая при этом вероятность ошибки P_{err}).

$$P_{err} = P_{lt} + P_{prc}, \quad P_{lt} \rightarrow \min, \quad P_{prc} \rightarrow \min. \quad (4.71)$$

С учетом предложенной модели системы фильтрации (подраздел 4.2) можно сделать вывод о том, что дополнительная оценка биспектральной плотности (сечение СП по коэффициенту асимметрии) для сегментов, содержащих только акустическое воздействие в силу симметричности закона распределения их мгновенных значений на длительном интервале квазистационарности будет обладать схожими свойствами с выявленными параметрами, полученными при анализе энергетических особенностей РС и областями пиковой стабильности биспектральной интенсивности для сегментов РС, содержащих полезную информационную составляющую [50]. В свою очередь биспектр обрабатываемого сегмента раскрывает аспект бикорреляции формантной структуры природы речеобразования – показывает степень взаимосвязи между формантами (рисунок 4.10), подтверждая асимметричность участков РС, содержащих полезную составляющую.

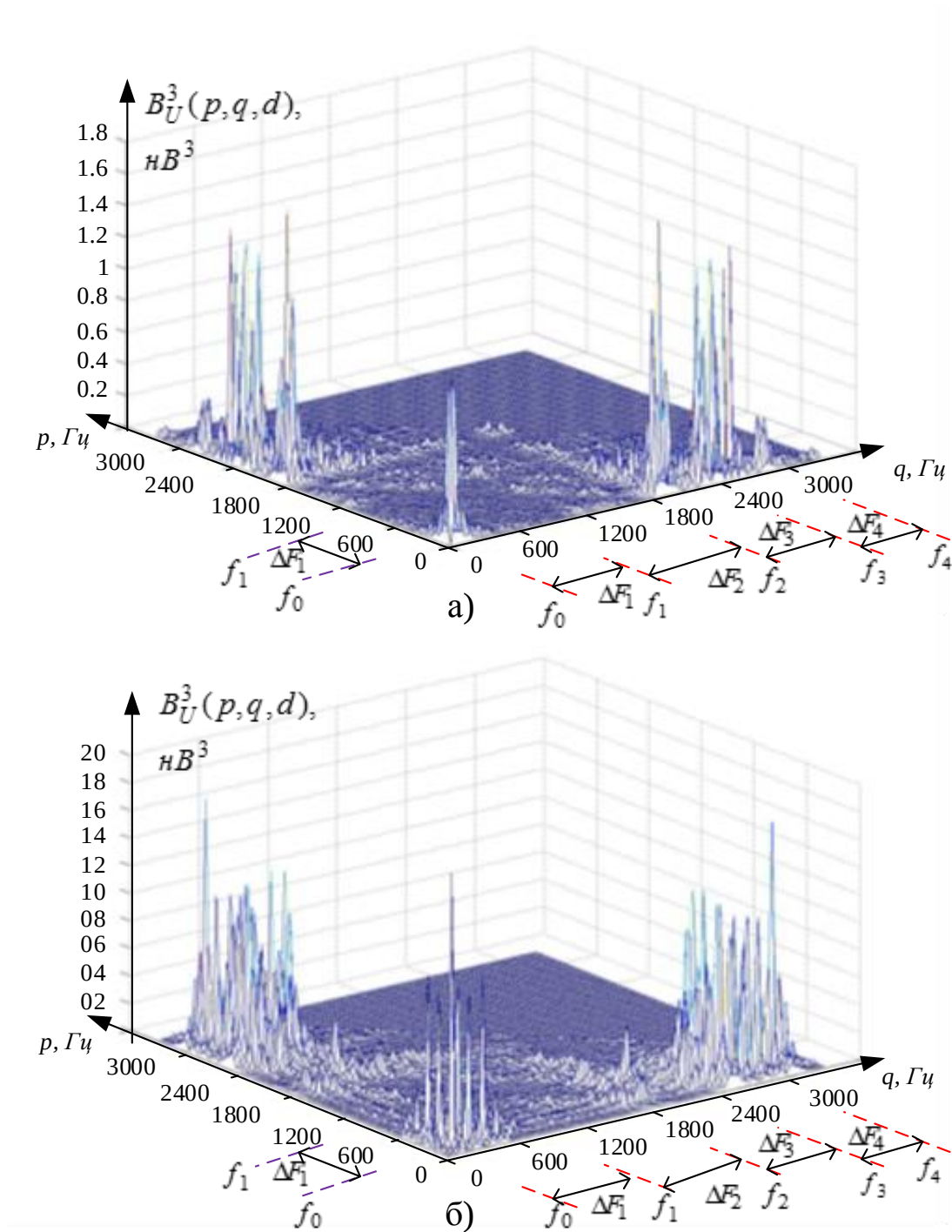


Рисунок 4.10 – Бикорреляция в обрабатываемом РС:

- а) биамплитуда сегмента РС (шум двигателя ОСШ 10 дБ), описывающего бикорреляционно сосредоточенный элемент речи – звук «К»;
- б) биамплитуда сегмента РС (КБГШ ОСШ 10 дБ), описывающего бикорреляционно сосредоточенный элемент речи – звук «О».

Исходя из вышеизложенного, необходимо определить совокупность параметров РС, получаемых путем анализа биспектральной плотности, называемые далее коэффициенты суммарной бикорреляции (КСБ) при $k = 1, 2, 3, \dots, \frac{F_d}{2}$:

$$\begin{aligned}\lambda_1(d) &= \|\Delta F_1, \Delta F_1\| = \sum_{p=f_0}^{p=f_1} \sum_{q=f_0}^{q=f_1} B_U^3(p, q, d), \\ \lambda_2(d) &= \|\Delta F_1, \Delta F_2\| = \sum_{p=f_0}^{p=f_1} \sum_{q=f_1}^{q=f_2} B_U^3(p, q, d), \\ \lambda_3(d) &= \|\Delta F_1, \Delta F_3\| = \sum_{p=f_0}^{p=f_1} \sum_{q=f_2}^{q=f_3} B_U^3(p, q, d), \\ \lambda_4(d) &= \|\Delta F_1, \Delta F_4\| = \sum_{p=f_0}^{p=f_1} \sum_{q=f_3}^{q=f_4} B_U^3(p, q, d),\end{aligned}\tag{4.72}$$

где $\lambda_1(d) - \lambda_4(d)$ – коэффициенты суммарной бикорреляции для сегмента d , $\|\Delta F_1, \Delta F_4\|$ – взаимосвязь формантных областей сегмента РС, $f_0 - f_4$ – границы формант в биспектральной плотности.

Анализ рисунка 4.10 и параметров РС позволяют определить основное правило выделения пауз, а именно определение участков РС, не содержащих полезной составляющей, посредством анализа асимметрии СП на длительных интервалах наблюдения (не более двух секунд). То есть оценка количества знакоперемен по каждому из КСБ (4.72 – 4.75) относительно их среднего значения на участках квазистационарности воздействия. С учетом первого требования, предъявляемого к процедуре выделения пауз, заключающегося в низкой вероятности «пропуска цели» в условиях воздействия на входе системы фильтрации РС, представленного только сигналами акустического воздействия, исследования в рамках предметной области показывают высокое количество знакоперемен по каждому из КСБ (таблица 4.2) для идеализированных и реальных акустических воздействий.

Таблица 4.2– Бикорреляционные параметры акустических воздействий

Вид шума	Параметр											
	β_1		β_2		β_3		β_4		β		P_{prc}	
	1 сек	2 сек	1 сек	2 сек	1 сек	2 сек	1 сек	2 сек	1 сек	2 сек	1 сек	2 сек
Шум автомагистрали	0.47	0.45	0.51	0.46	0.48	0.47	0.45	0.47	0.54	0.46	0.01	0.01
Шум двигателя	0.41	0.40	0.38	0.37	0.35	0.35	0.36	0.35	0.36	0.40	0.01	0.01
Шум ветра	0.36	0.36	0.37	0.37	0.35	0.37	0.35	0.35	0.36	0.36	0.01	0.02
Шум города	0.41	0.39	0.40	0.39	0.36	0.38	0.35	0.36	0.40	0.39	0.01	0.01
Шум пламени	0.45	0.47	0.45	0.43	0.43	0.43	0.44	0.43	0.51	0.47	0.01	0.02
Шум боя	0.35	0.36	0.35	0.36	0.35	0.35	0.35	0.36	0.35	0.37	0.02	0.03
Шум вертолета	0.37	0.38	0.37	0.37	0.37	0.37	0.37	0.37	0.37	0.36	0.02	0.03
Шум поезда	0.33	0.35	0.33	0.35	0.33	0.36	0.35	0.34	0.32	0.35	0.03	0.03
Тоновой шум	0.50	0.49	0.49	0.49	0.51	0.49	0.49	0.49	0.51	0.48	0	0

* результаты, представленные в таблице, получены при частоте дискретизации 8000 Гц и сегменте 32 мс (256 отсчетов) имеют высокую коррелированность с испытаниями на других частотах дискретизации и других отрезках локальной стационарности (не превышающих 40 мс).

Рассмотрим выделение бикорреляционных параметров акустических воздействий:

$$\beta_j = \frac{1}{D} \cdot \sum_{d=1}^{d=D \pm} \mu_j^{\pm}(d), \quad (4.73)$$

где $\mu_j^{\pm}(d)$ – признак смены полярности значения j -го КСБ сегмента относительно его среднего значения в рамках интервала наблюдения при $d = 1, 2, 3, \dots, D$ определяемый согласно следующих выражений:

$$\begin{aligned}
\lambda_j(d) < \lambda_j^a \ \& \ \lambda_j(d+1) > \lambda_j^a \Rightarrow \mu_j^\pm(d) = 1, \\
\lambda_j(d) > \lambda_j^a \ \& \ \lambda_j(d+1) < \lambda_j^a \Rightarrow \mu_j^\pm(d) = 1, \\
\lambda_j(d) \leq \lambda_j^a \ \& \ \lambda_j(d+1) \leq \lambda_j^a \Rightarrow \mu_j^\pm(d) = 0, \\
\lambda_j(d) \geq \lambda_j^a \ \& \ \lambda_j(d+1) \geq \lambda_j^a \Rightarrow \mu_j^\pm(d) = 0,
\end{aligned} \tag{4.74}$$

где $\lambda_j(d)$ – значение j -го КСБ для сегмента d ; $\lambda_j^a(d)$ – среднее значение j -го КСБ сегмента в рамках интервала наблюдения при $d = 1, 2, 3, \dots, D$ определяемое согласно следующего выражения:

$$\lambda_j^a = \frac{1}{D} \cdot \sum_{d=1}^{d=D} \lambda_j(d). \tag{4.75}$$

Анализ (табл. 4.2) и (рисунка 4.11) позволяет выявить следующее правило выделения длительных участков паузы в РС:

$$\begin{cases} \beta_j^D \geq \beta_{nbn}^o \Rightarrow J_D \triangleright 1, \\ \beta_j^D \geq \beta_{wbn}^o \Rightarrow J_D \triangleright 1, \end{cases} \tag{4.76}$$

где β_j^D – совокупность значений количества знакоперемен по j -м КСБ при анализе D – выборки сегментов РС; J_D – районная оценка функции признака паузы; β_{nbn}^o , β_{wbn}^o – пороговые значения коэффициентов симметрии для узкополосного и КБГШ соответственно.

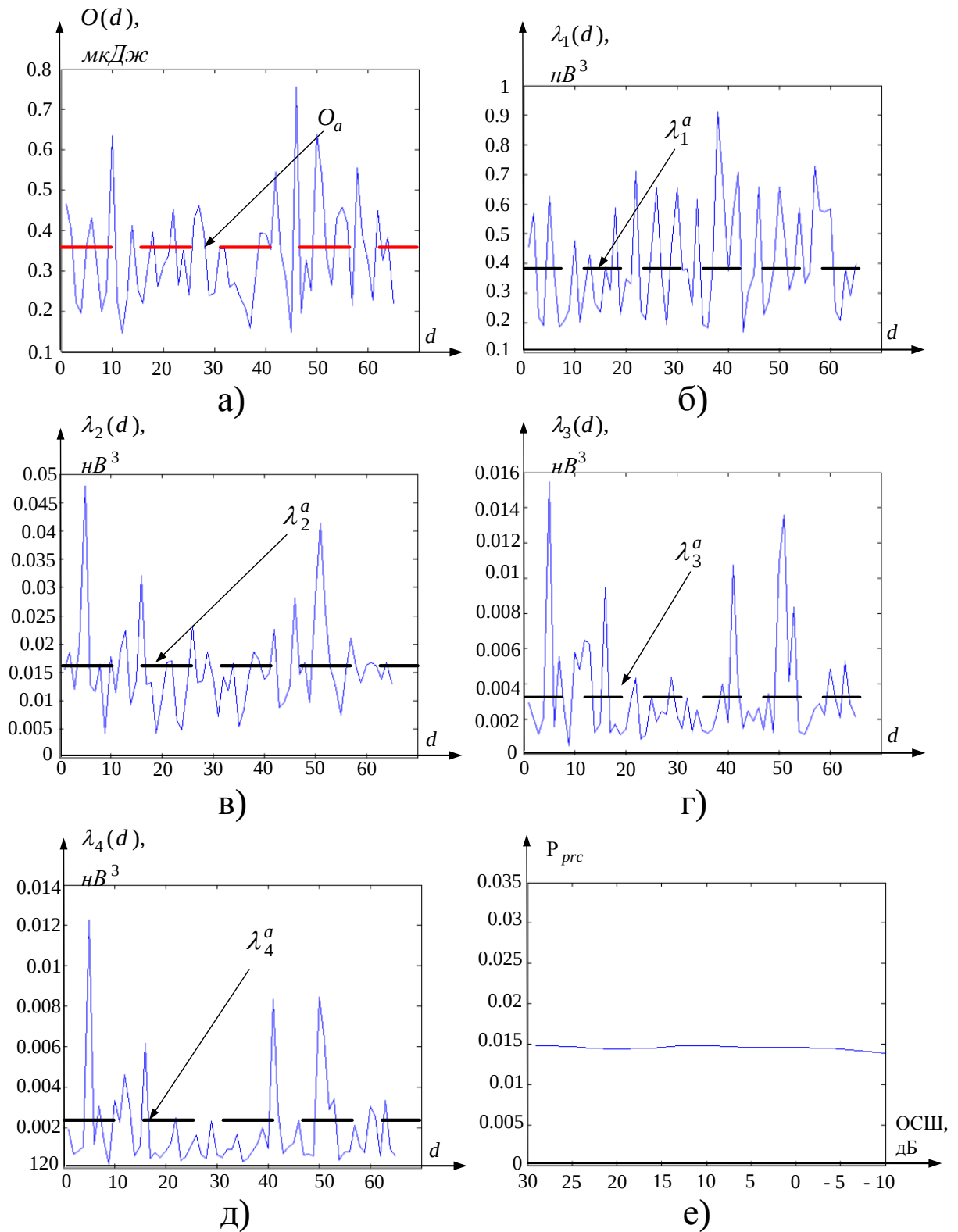


Рисунок 4.11 – Бикорреляционные свойства реальных акустических воздействий:
а) значение мгновенной энергии посегментно; б) значение первого КСБ посегментно;
в) значение второго КСБ посегментно; г) значение третьего КСБ посегментно;
д) значение четвертого КСБ посегментно; е) зависимость вероятности «пропуска цели» от ОСШ на входе системы фильтрации.

Необходимо отметить, что значение коэффициента симметрии [11, 70] лежит в диапазоне от 0 до 0,5. Исходя из результатов исследований определим $\beta_{nbn}^o = 0,35$; $\beta_{wbn}^o = 0,45$. Как следует из табл. 4.2 на длительных интервалах наблюдения, не превышающих 2 секунд, как реальные, так и идеализированные акустические воздействия стремятся к гауссовскому закону распределения мгновенных значений анализируемого сигнала. При этом количество знакоперемен стремилось к коэффициенту 0.5 от общего количества отсчетов, что характеризуется высокой симметричностью функции распределения. Вследствие этого для оценки биспектральных коэффициентов в вопросах оценки асимметрии на вход системы анализа подавались последовательности отсчетов длительностью 1 и 2 секунды различных реализаций (по 10 реализаций), указанных выше типов шумов, далее посегментно оценивались коэффициенты β_j количества знакоперемен относительно среднего для каждого КСБ (табл. 4.2). Анализ данной таблицы (для задачи минимизации пропуска цели для всех типов шумов) показывает, что выбор $\beta_{nbn}^o = 0,35$, $\beta_{wbn}^o = 0,45$ обеспечивает низкую вероятность пропуска цели для всех типов и реализаций акустических воздействий в случае подачи на вход системы анализа отсчетов сигнала содержащего только акустическое воздействие. Вышеизложенные факты позволяют сделать вывод, о том, что применение такого подхода в вопросе выделения длительных пауз РС, представленных интенсивным акустическим воздействием, позволяет добиться вероятности "пропуска цели" не более 0,015.

Следовательно, вопрос выделения паузы в условиях, когда на вход системы фильтрации поступают различные последовательности сегментов локальной стационарности, представляющих как активные участки, так и паузы, лежит в плоскости свойств асимметрии РС на длительном интервале наблюдения путем оценки дисбаланса параметров, описывающих все анализируемые сегменты и чем больше разница, тем меньше количество знакоперемен по каждому КСБ. Факт преимущества биспектральных оценок относительно "классического" спектрального анализа проиллюстрирован на (рисунке 4.12).

Исходя из вышеизложенного, необходимо определить параметры сигнала, приведенных для сравнения с КСБ и получаемых путем анализа спектральной плотности мощности, называемые далее коэффициенты районной мощности:

$$O_1(d) = A_U^2 \{\Delta F_1\} = \sum_{k=f_0}^{k=f_1} A_U^2(k, d), \quad (4.77)$$

$$O_2(d) = A_U^2 \{\Delta F_2\} = \sum_{k=f_1}^{k=f_2} A_U^2(k, d), \quad (4.78)$$

$$O_3(d) = A_U^2 \{\Delta F_3\} = \sum_{k=f_2}^{k=f_3} A_U^2(k, d), \quad (4.79)$$

$$O_4(d) = A_U^2 \{\Delta F_4\} = \sum_{k=f_3}^{k=f_4} A_U^2(k, d), \quad (4.80)$$

где $O_1(d) - O_4(d)$ – коэффициенты районной мощности для сегмента d ; $\Delta F_1 - \Delta F_4$ – ограниченные области спектральной плотности мощности, $A_U^2(k)$ – оценка спектра мощности; $f_0 - f_4$ – границы спектральных областей.

Рисунок 4.12 указывает на преимущество КСБ относительно коэффициентов районной мощности, т.е. при подаче на вход процедуры оценки количества знакоперемен сигнала, содержащего как полезную составляющую, так и акустическое воздействие, значения данных коэффициентов для каждого коэффициенты суммарной бикорреляции стремятся к 0, что связано со значением среднего при учете пиковых значений на активных участках. Далее на основе алгоритма выкалывания максимального значения среднее переоценивается (снижается) и опять оценивается относительное количество знакоперемен [54, 56]. Алгоритм выкалывания функционирует до момента достижения относительного количества знакоперемен порогового значения $\beta_{nbn}^o = 0,35$, $\beta_{wbn}^o = 0,45$, оставшиеся сегменты считаются сегментами паузы, а максимальное значение является пороговым значением для данной реализации анализируемого РС.

В качестве сравнительного параметра показана степень различия максимального уровня коэффициентов для активных участков РС и максимального значения для паузы.

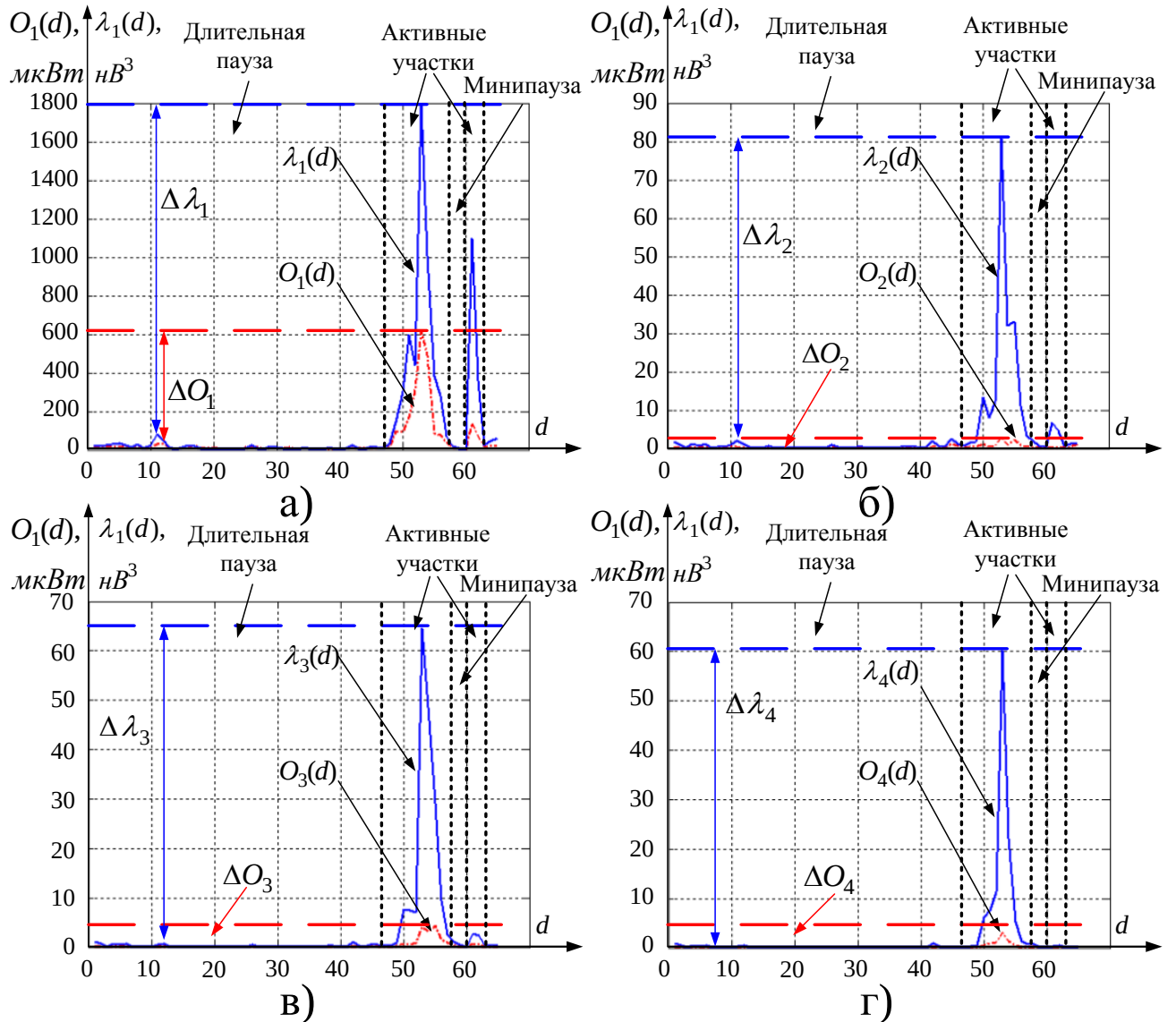


Рисунок 4.12 – Сравнительный анализ КСБ и параметров, полученных путем анализа оценок спектра мощности для совокупности сегментов локальной стационарности РС в течении интервала наблюдения (среднее ОСШ 0 дБ – шум боя):

а) значения коэффициентов суммарной бикорреляции и районной мощности для первой частотной полосы; б) значения коэффициентов суммарной бикорреляции и районной мощности для второй частотной полосы; в) значения коэффициентов суммарной бикорреляции и районной мощности для третьей частотной полосы; г) значения коэффициентов суммарной бикорреляции и районной мощности для четвертой частотной полосы.

Исходя из анализа рисунка 4.12, можно сделать вывод, что учет КСБ позволяет выделить участки пауз более точно, используя дисбаланс значений данных коэффициентов, т.е. основой выделения участков, не содержащих полезной составляющей, является информация о пороговых значениях по каждому КСБ, полученных как максимальные для совокупности сегментов, содержащих только акустическое

воздействие. Следовательно, определение пороговых значений сопряжено с подсчетом количества знакоперемен по каждому КСБ для различных последовательностей сегментов анализа. Исходя из данного факта можно определить следующую процедуру, которая основана на итеративном алгоритме «с выкалыванием» максимального значения, вследствие чего последовательность значений для каждого КСБ будет итеративно уменьшаться пока количество знакоперемен (4.74) не превысит пороговое значение (4.77, 4.78). Для эффективной работы процедуры выделения паузы необходимо дополнительно решить вопрос его адаптации к условиям зашумления только КБГШ, так как в реальных условиях работы УОРС это может быть следствием, например, неисправного микрофона.

Исходя из свойства КБГШ, заключающегося в равномерности оценки спектральной плотности мощности, определено следующее:

$$R_1(d) = \frac{1}{F_b} \cdot \sum_{k=0}^{k=F_b} A_U^2(k, d), \quad (4.81)$$

$$R_2(d) = \frac{1}{F_a} \cdot \sum_{k=0}^{k=F_a} A_U^2(k, d), \quad (4.82)$$

$$R_3(d) = \frac{1}{F_b - F_a} \cdot \sum_{k=F_a}^{k=F_b} A_U^2(k, d), \quad (4.83)$$

$$\frac{R_1(d)}{R_2(d)} \rightarrow 1; \frac{R_1(d)}{R_3(d)} \rightarrow 1; \frac{R_2(d)}{R_3(d)} \rightarrow 1 \quad (4.84)$$

$$\left\{ \begin{array}{l} \frac{R_1(d)}{R_2(d)} \leq r^o; \\ \frac{R_1(d)}{R_3(d)} \leq r^o; \\ \frac{R_2(d)}{R_3(d)} \leq r^o; \end{array} \right. \Rightarrow H(d) = 1, \quad (4.85)$$

где $H(d)$ – значение бинарной функции признака зашумления КБГШ; $R_1(d)$ – коэффициент среднего значения спектральных компонент во всей анализируемой полосе частот для сегмента d ; $R_2(d)$ – коэффициент среднего значения спектральных компонент в полосе низких частот для сегмента d ; $R_3(d)$ – коэффициент среднего значения спектральных компонент в полосе высоких частот для сегмента d ; r^o – пороговый относительный коэффициент.

Исходя из проведенных исследований [56] $r^o = 1,5$; F_a, F_b – граничные частоты, определяемые согласно:

$$F_a = \frac{F_d}{4}, F_b = \frac{F_d}{2}, \quad (4.86)$$

где F_d – нормированная частота дискретизации.

Результативность разработанного алгоритма выделения сегментов паузы при обработке РС проиллюстрирована на (рисунке 4.13).

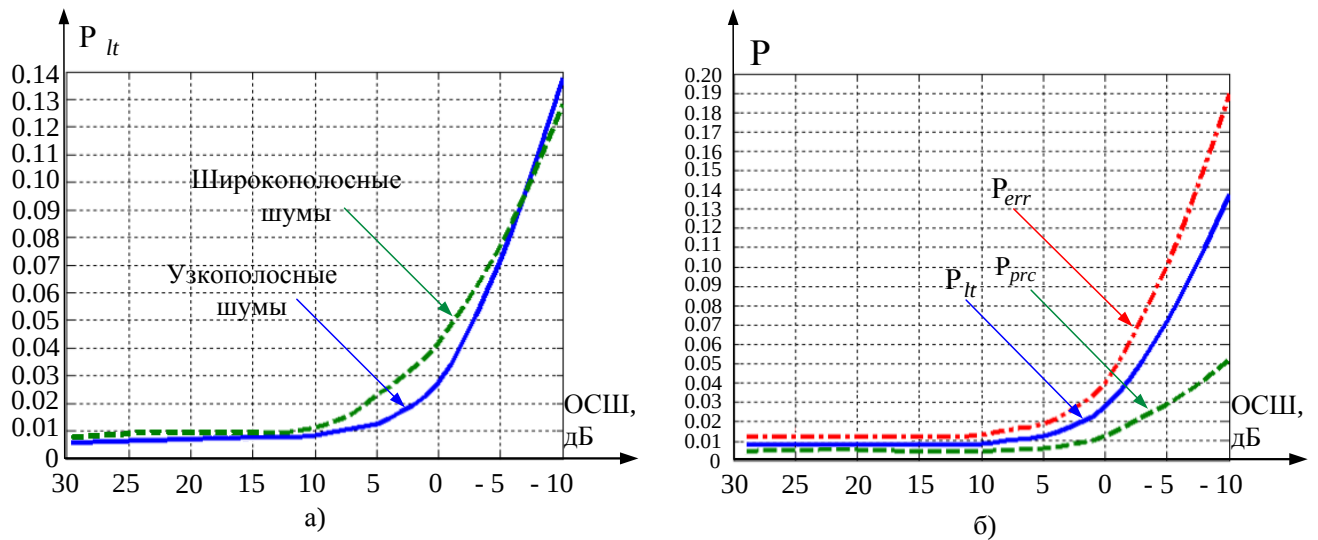


Рисунок 4.13–Результативность процедуры выделения паузы в РС в УИАВ: а) зависимость вероятности «ложной тревоги» от ОСШ для широкополосных и узкополосных акустических воздействий; б) зависимость вероятности ошибочного выделения сегментов паузы от ОСШ (на примере шума города).

Анализ (рисунка 4.13) показывает, что всегда существует вероятность ложного выделения сегментов паузы, кроме того, данная вероятность имеет тенденцию к росту с увеличением энергетики воздействия. Данные факты обусловлены как некоторой нестабильностью самих шумов, так и наличием в обрабатываемом РС неречевых конструкций или переходных элементов, не имеющих четкой формантной структуры. Также стоит отметить, что выделение паузы в РС не решает полностью вопрос оценки спектральных характеристик акустического воздействия в силу наличия "атипичных выбросов" как в самих реализациях шума, так вариативностью полезного сигнала в различных по типу паузах.

Исходя из проведенного анализа параметрических особенностей идеализированных и реальных акустических воздействий для стабилизации усредненной

оценки спектральных характеристик акустического воздействия предложена процедура, основанная на коэффициенте стабильности эффективной полосы частот $\Delta P(d)$ акустического воздействия α :

$$\alpha = \frac{\max[\Delta P(d)]}{\Delta P_a}, \quad (4.87)$$

где d – последовательный номер сегмента обработки в рамках интервала наблюдения при $d = 1, 2, 3, \dots, D$; ΔP_a – среднее значение эффективной полосы частот акустического воздействия на квазистационарном длительном интервале наблюдения, определяемое согласно следующего выражения:

$$\Delta P_a = \frac{1}{M} \cdot \sum_{d=1}^{d=M} \Delta P(d). \quad (4.88)$$

где M – количество выделенных сегментов паузы РС при $d = 1, 2, 3, \dots, D$.

Исходя из вышеизложенного можно определить правило отбора выделенных сегментов паузы для определения усредненной оценки спектральных характеристик акустического воздействия: $\alpha \leq \alpha_{nbn}^o$, где $\alpha_{nbn}^o = 2$ – пороговое значение стабильности эффективной полосы частот узкополосных акустических воздействий.

В случае воздействия КБГШ правило выбора сегментов в силу равномерности спектральной плотности мощности данного вида воздействий носит характер интервальной оценки:

$$\Delta P_a^{wbn} - \xi < \Delta P(d) \leq \Delta P_a^{wbn} + \xi, \quad (4.89)$$

где ξ – параметр доверительного интервала оценки средней эффективной полосы КБГШ ΔP_a^{wbn} , определяемой согласно следующего выражения:

$$\Delta P_a^{wbn} = 0.9 \cdot \frac{F_d}{2}, \quad (4.90)$$

где F_d – нормированная частота дискретизации.

Необходимо отметить, что в случае узкополосных акустических воздействий процедура стабилизации выполнена на основе итерационного алгоритма «с выкалыванием» максимального значения с дальнейшей переоценкой параметров, а для ситуации воздействия КБГШ на базе алгоритма «разового отсеивания».

После выделения сегментов паузы бинарная функция признака паузы $J(d)$ имеет значение «0» для активных участков и «1» для паузы, однако после процедуры стабилизации оценки спектральных характеристик акустического воздействия на основе данной функции мы получаем бинарную функцию усреднения спектральных характеристик сегментов паузы $J_P(d)$, имеющей значение «0» для активных участков и «отбракованных» сегментов паузы в остальных случаях «1».

Далее формируется массив значений спектральных данных $A_U^a(k, d)$ согласно следующему выражению:

$$A_U^a(k, d) = A_U(k, d) \cdot J_P(d). \quad (4.91)$$

Необходимо отметить, что для сегментов, которые не определены для усреднения спектральных характеристик акустического воздействия данный массив имеет значения «0» для всех спектральных компонент.

На основании полученного массива спектральных данных (4.91) производится усреднение спектральных характеристик сегментов обработки на длительном интервале наблюдения:

$$\langle A_E(k, d) \rangle = \frac{1}{\sum_{d=1}^{d=D} J_P(d)} \cdot \sum_{d=1}^{d=D} A_U^a(k, d), \quad (4.92)$$

где $\langle A_E(k, d) \rangle$ – усредненная оценка спектральных характеристик акустического воздействия для сегмента d .

Результативность процедуры стабилизации и непосредственно самой оценки усредненных спектральных характеристик акустического воздействия проиллюстрирована на (рисунке 4.14). На рисунках 4.14 (в, г) показан РС во временной области и усредненная оценка амплитудного спектра акустического воздействия для

случая присутствия РС соответственно, а на рисунках 4.14 (а, б) представлены оценки данного акустического шумового воздействия при отсутствии РС.

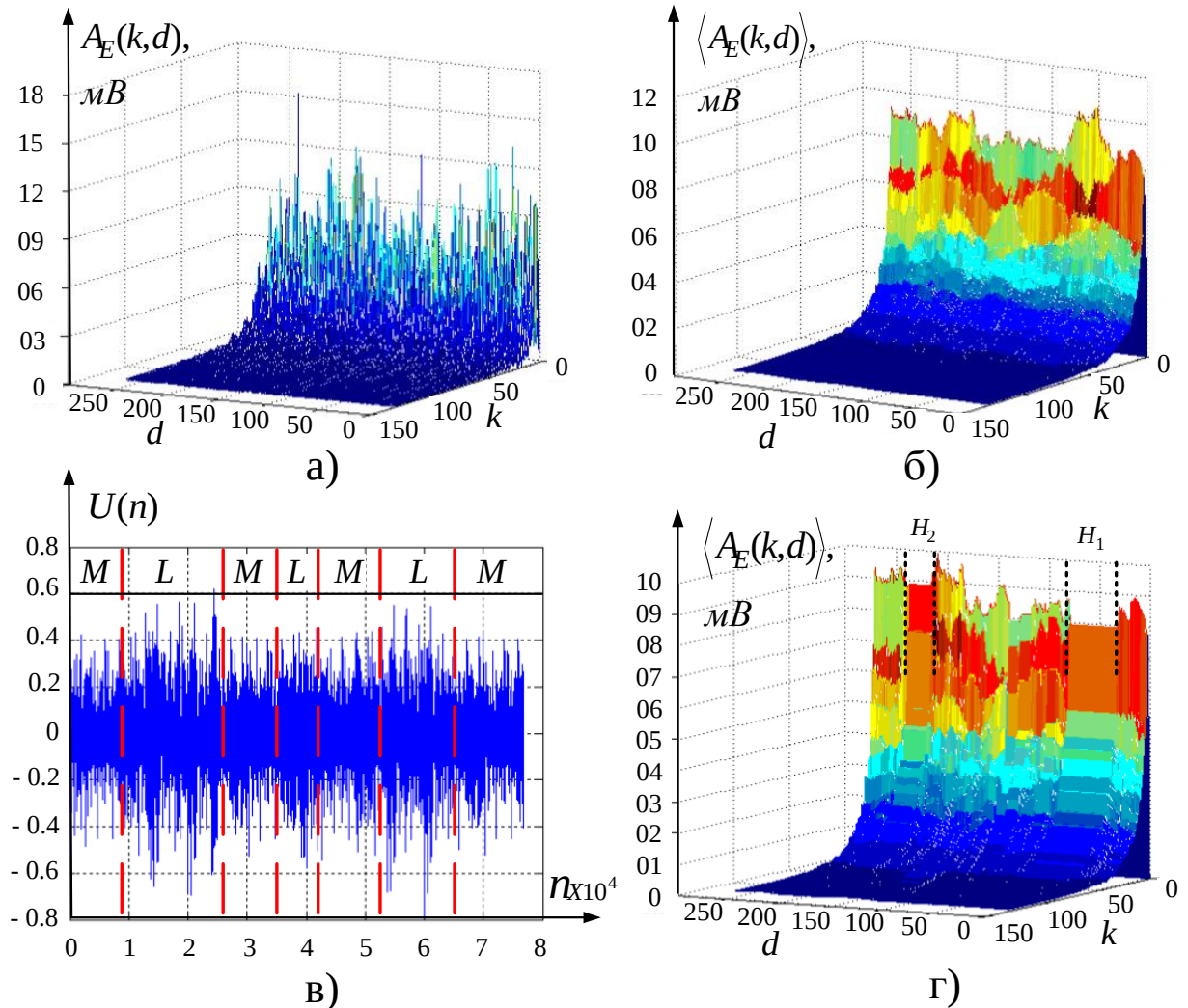


Рисунок 4.14 – Результативность процедуры оценивания спектральных характеристик акустического воздействия шум города ОСШ = – 10 дБ): а) истинная оценка спектра амплитуд акустического воздействия посегментно; б) усредненная оценка спектра амплитуд акустического воздействия посегментно для случая отсутствия полезного сигнала (высокая степень корреляции с истинной оценкой); в) реализация РС во временной области на длительности 10 секунд (участки M – паузы, L – активные участки); г) усредненная оценка спектра амплитуд акустического воздействия посегментно для случая присутствия полезного сигнала (участки H_1, H_2 – сохранение усредненной оценки спектральных характеристик акустического воздействия).

Выводы по четвертому разделу

1. Исследования основных свойств РС позволяют сделать вывод о том, что использование предложенной адаптивной системы фильтрации на основе спектрального вычитания, способной изменять свои параметры, учитывая характеристики акустических воздействий, позволяет значительно повысить качество обрабатываемого РС.

2. Существует возможность повышения степени адаптации системы фильтрации РС к характеристикам реальных акустических воздействий на длительных интервалах квазистационарности за счет полиспектрального анализа РС.

3. Предложенная математическая модель системы фильтрации РС на основе полиспектрального анализа основана на следующих принципах:

- для усовершенствования выбран класс систем фильтрации на основе спектрального вычитания;
- параметрическая адаптация системы фильтрации осуществляется по входному сигналу путем оценивания характеристик акустических воздействий на длительном интервале квазистационарности с использованием результатов оценок мультиспектров.

4. Разработанная математическая модель системы фильтрации РС на основе полиспектрального анализа устраняет большинство недостатков существующих подходов к шумоподавлению и позволяет разработать алгоритмы анализа и обработки РС при наличии шумовых акустических воздействий.

5. Разработанная математическая модель позволяет определить степень влияния параметров адаптации на показатели ОСШ и субъективную оценку качества РС на выходе системы фильтрации.

6. На основе дополнительных условий и ограничений к предложенной математической модели доказана и обоснована ее полнота для задачи разработки алгоритмов адаптивной цифровой фильтрации речевого сигнала в УИАВ.

5. РАЗРАБОТКА СПОСОБОВ РЕАЛИЗАЦИИ ПРОЦЕДУР ФИЛЬТРАЦИИ РЕЧЕВОГО СИГНАЛА В ЗАДАЧЕ ШУМОПОДАВЛЕНИЯ

5.1 Вводные замечания

На основании разработанной математической модели (раздел 4.2), которая отличается дополнительным введением оценки биспектральной плотности относительно существующих систем фильтрации с использованием спектрального вычитания, разработан метод адаптивной цифровой фильтрации речевого сигнала с целью шумоподавления. Он отличается от известных тем, что учитывает бикорреляционные свойства сигнала на основе анализа мультиспектров с применением полученных параметров в процедурах выделения сегментов паузы, стабилизации и оценки спектральных характеристик акустического воздействия, которые включены в его состав. Разработанный метод не зависит от используемой частоты дискретизации и не требует априорных сведений об ОСШ и характере акустического воздействия.

В данном разделе рассмотрен предложенный метод адаптивной цифровой фильтрации речевого сигнала в условиях акустического зашумления (подраздел 5.2), проведено обоснование оценки критерия качества речевого сигнала на выходе системы обработки (подраздел 5.3), показана оценка повышения качества речевого сигнала при шумоподавлении согласно предлагаемым техническим решениям (подраздел 5.4).

5.2 Метод адаптивной фильтрации речевого сигнала в условиях действия акустических помех

В разработанном методе используется адаптация к особенностям и параметрам акустического воздействия (для каждого обрабатываемого сегмента используется "индивидуальная" усредненная оценка спектральных характеристик акустического воздействия) и имеет широкий динамический диапазон входных значений

ОСШ от 35 до -10 дБ. Его составным элементом является алгоритм адаптивной цифровой фильтрации РС в УИАВ, который проиллюстрирован в виде блок-схемы на рисунке 5.1. Для раскрытия вопросов адаптации необходимо провести анализ блок-схемы алгоритма через подробную последовательность элементарных операций.

Этапы алгоритма:

1) Ввод последовательности отсчетов $U(n, d)$ сегмента обработки d , обнуление счетчика количества итераций i .

2) Условие проверки наличия необходимого количества сегментов для оценки спектральных характеристик акустического воздействия:

$$d \geq D_m, \quad (5.1)$$

где d – последовательный номер сегмента, так что d – текущий сегмент обработки, $d-1$ – предшествующий текущему сегменту обработки и т.д., D_m – суммарное количество сегментов, необходимое для состоятельной оценки спектральных характеристик акустического воздействия (для частоты дискретизации 8 000 Гц и сегмента локальной стационарности 32 мс – 256 отсчетов $D_m = 64 \Leftrightarrow 2 \text{сек} ; 32 \Leftrightarrow 1 \text{сек}$).

3) {Невыполнение условия (5.1)} оценка комплексного спектра Фурье обрабатываемого сегмента d .

4) Оценка спектра мощности $A_U^2(k, d)$ и биамплитуды $B_U^3(p, q, d)$ обрабатываемого сегмента d .

5) Оценка коэффициентов суммарной бикорреляции (4.72–4.75).

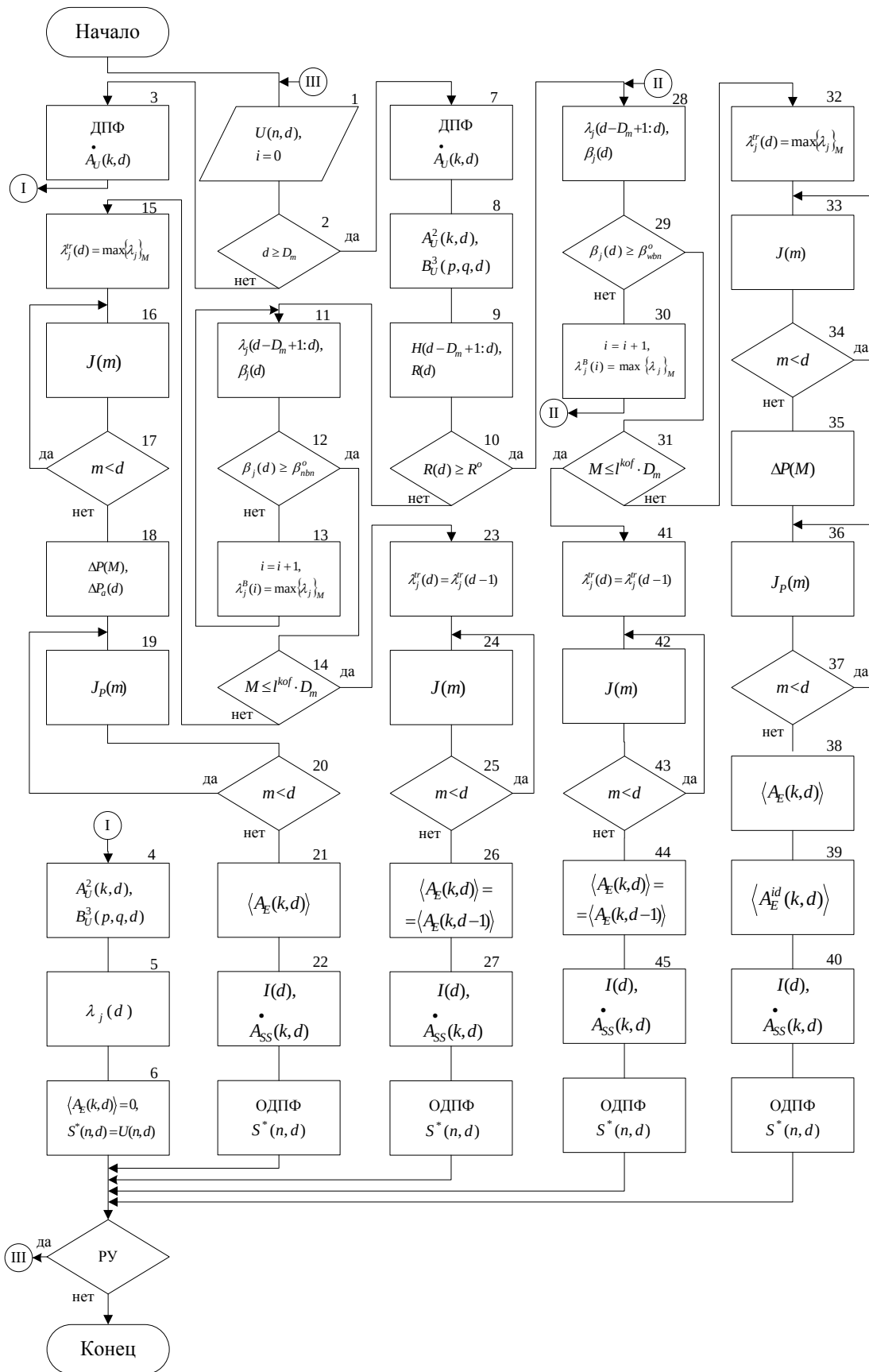


Рисунок 5.1 – Блок схема алгоритма адаптивной цифровой фильтрации РС в условиях акустического зашумления, реализующего заявленный метод

6) Получение значений сигнала на выходе системы фильтрации $S^*(n, d)$:

$$S^*(n, d) = U(n, d), \quad (5.2)$$

где $U(n, d)$ – сигнал на входе системы фильтрации.

Запись нулевой усредненной оценки спектральных характеристик акустического воздействия $\langle A_E(k, d) \rangle \forall k$:

$$\langle A_E(k, d) \rangle = 0. \quad (5.3)$$

7) {Выполнение условия (5.1)} аналогично пункта 3.

8) Аналогично пункта 4.

9) Оценка коэффициентов среднего значения спектральных компонент $R_1(m) - R_3(m)$ и бинарной функции признака зашумления КБГШ $H(m)$ (4.81 – 4.85), оценка признака зашумления КБГШ $R(d)$ по следующему выражению:

$$R(d) = \sum_{m=d-D_m+1}^{m=d} H(m). \quad (5.4)$$

10) Условие проверки наличия зашумления обрабатываемого РС КБГШ:

$$R(d) \geq R^o, \quad (5.5)$$

где $R^o = \frac{D_m}{3}$ – идентификатор КБГШ.

11) {Невыполнение условия (5.5)} оценка КСБ $\lambda_j(d)$ и количества знаков перемен при $\beta_j(d) \Big|_{M=D_m}$:

$$\lambda_1(d) = \|\Delta F_1, \Delta F_1\| = \sum_{p=f_0}^{p=f_1} \sum_{q=f_0}^{p=f_1} B_U(p, q, d), \quad (5.6)$$

$$\lambda_2(d) = \|\Delta F_1, \Delta F_2\| = \sum_{p=f_0}^{p=f_1} \sum_{q=f_1}^{p=f_2} B_U(p, q, d), \quad (5.7)$$

$$\lambda_3(d) = \|\Delta F_1, \Delta F_3\| = \sum_{p=f_0}^{p=f_1} \sum_{q=f_2}^{p=f_3} B_U(p, q, d), \quad (5.8)$$

$$\lambda_4(d) = \|\Delta F_1, \Delta F_4\| = \sum_{p=f_0}^{p=f_1} \sum_{q=f_3}^{q=f_4} B_U(p, q, d), \quad (5.9)$$

где $\lambda_1(d) - \lambda_4(d)$ – коэффициенты суммарной бикорреляции для сегмента d ,
 $\|\Delta F_1, \Delta F_4\|$ – взаимосвязь формантных областей сегмента РС,
 $f_0 - f_4$ – границы формант в биспектральной плотности.

$$\beta_j(d) \Big|_{M=D_m} = \frac{1}{D_m} \cdot \sum_{\Omega M}^{\pm} \mu_j^{\pm}(m), \quad (5.10)$$

где $\mu_j^{\pm}(m)$ – признак смены полярности значения j -го КСБ сегмента относительно его среднего значения в рамках интервала наблюдения при $m = d - D_m + 1, \dots, d$ определяемый согласно следующего выражения:

$$\begin{aligned} \lambda_j(m) < \lambda_j^a(d) \Big|_{M=D_m} \quad \& \quad \lambda_j(m+1) \geq \lambda_j^a(d) \Big|_{M=D_m} \Rightarrow \mu_j^{\pm}(m) = 1, \\ \lambda_j(m) \geq \lambda_j^a(d) \Big|_{M=D_m} \quad \& \quad \lambda_j(m+1) < \lambda_j^a(d) \Big|_{M=D_m} \Rightarrow \mu_j^{\pm}(m) = 1, \\ \lambda_j(m) < \lambda_j^a(d) \Big|_{M=D_m} \quad \& \quad \lambda_j(m+1) < \lambda_j^a(d) \Big|_{M=D_m} \Rightarrow \mu_j^{\pm}(m) = 0, \\ \lambda_j(m) > \lambda_j^a(d) \Big|_{M=D_m} \quad \& \quad \lambda_j(m+1) > \lambda_j^a(d) \Big|_{M=D_m} \Rightarrow \mu_j^{\pm}(m) = 0, \end{aligned} \quad (5.11)$$

где $\lambda_j(m)$ – значение j -го КСБ для сегмента m ; $\lambda_j^a(d)$ – среднее значение j -го КСБ сегмента в рамках интервала наблюдения при $m = d - D_m + 1, \dots, d$ определяемое согласно следующего выражения:

$$\lambda_j^a(d) \Big|_{M=D_m} = \frac{1}{M} \cdot \sum_{\Omega M} \lambda_j(m). \quad (5.12)$$

12) Проверка выполнения условия по пороговому значению для узкополосного воздействия j -му КСБ:

$$\beta_j(d) \Big|_{M=D_m} \geq \beta_{nbn}^o \quad (5.13)$$

где β_{nbn}^o – пороговое значение количества знакоперемен для узкополосного акустического воздействия.

13) Невыполнение условия (5.13)} по j -му КСБ выполнение алгоритма "с выкалыванием" максимального значения:

$$i = i + 1, \quad (5.14)$$

$$\lambda_j^B(i) = \max\{\lambda_j\}_M \quad (5.15)$$

$$M = M - \lambda_j^B(i), \quad (5.16)$$

$$\lambda_j^a(d) \Big|_{M=D_m-i} = \frac{1}{M} \cdot \sum_{\Omega M} \lambda_j(m), \quad (5.17)$$

$$\begin{aligned} \lambda_j(m) < \lambda_j^a(d) \Big|_{M=D_m-i} \ \& \ \lambda_j(m+1) \geq \lambda_j^a(d) \Big|_{M=D_m-i} \Rightarrow \mu_j^\pm(m) = 1, \\ \lambda_j(m) \geq \lambda_j^a(d) \Big|_{M=D_m-i} \ \& \ \lambda_j(m+1) < \lambda_j^a(d) \Big|_{M=D_m-i} \Rightarrow \mu_j^\pm(m) = 1, \\ \lambda_j(m) < \lambda_j^a(d) \Big|_{M=D_m-i} \ \& \ \lambda_j(m+1) < \lambda_j^a(d) \Big|_{M=D_m-i} \Rightarrow \mu_j^\pm(m) = 0, \\ \lambda_j(m) > \lambda_j^a(d) \Big|_{M=D_m-i} \ \& \ \lambda_j(m+1) > \lambda_j^a(d) \Big|_{M=D_m-i} \Rightarrow \mu_j^\pm(m) = 0, \end{aligned} \quad (5.18)$$

$$\beta_j(d) \Big|_{M=D_m-i} = \frac{1}{M} \cdot \sum_{\Omega M} \mu_j^\pm(m), \quad (5.19)$$

где $\lambda_j^B(i)$ – максимальное значение анализируемого массива M на i -й итерации для j -го КСБ.

Необходимо отметить, что реализация алгоритма "с выкалыванием" может быть осуществлена и с остановкой по количеству итераций, которая в общем случае определяет "жесткость" правила усреднения массива со спектральными характеристиками сегментов, определенных как пауза, т.е. в случае выполнения условия (5.13):

14) {Выполнение условия (5.13)} по *всем* КСБ дополнительная проверка наличия количества сегментов для состоятельной оценки спектральных характеристик акустического воздействия:

$$M \leq l^{kof} \cdot D_m, \quad (5.20)$$

где l^{kof} – коэффициент "жесткости" правила переоценки, т.е. при $l^{kof} = 1$ переоценивание спектральных характеристик и параметров биспектральной области происходит только при выделении длительной паузы на протяжении всего интервала наблюдения – "жесткое правило", или при $l^{kof} = 0,5$ – "мягкое правило", т.е. позволяет производить оценку спектральных характеристик, используя не только длительные паузы, но и "минипаузы", однако повышает риск попадания в оценку переходных элементов РС. На (рис. 5.2 г) показана результативность процедуры оценки при $l^{kof} = 0,75$, – на участке между H_1 и H_2 происходит постоянное переоценивание спектральных характеристик воздействия без наличия в обрабатываемой последовательности сегментов длительной паузы – рисунок 5.2 в).

15) {Невыполнение условия (5.20)} оценка пороговых значений для КСБ:

$$\lambda_j^{tr}(d) = \max\{\lambda_j\}_M \quad (5.21)$$

где $\lambda_j^{tr}(d)$ – пороговое значение для j -го КСБ для сегмента d .

16-17) Оценка бинарной функции признака паузы $J(m)$ для всей последовательности обрабатываемых сегментов на длительном интервале наблюдения при $m = d - D_m + 1, \dots, d$:

$$\begin{aligned} \lambda_j(m) \leq \lambda_j^{tr}(d) &\Rightarrow J(m) = 1, \\ \lambda_j(m) > \lambda_j^{tr}(d) &\Rightarrow J(m) = 0. \end{aligned} \quad (5.22)$$

18) Оценка для каждого сегмента, определенного как пауза, эффективной полосы частот $\Delta P|_M$, в которой сосредоточено 90% мощности, и коэффициента стабильности $\alpha(d)$ для сегмента d :

$$\alpha(d) = \frac{\max[\Delta P]_M}{\Delta P_a|_M}, \quad (5.23)$$

где d – последовательный номер сегмента обработки в рамках интервала наблюдения при $m = d - D_m + 1, \dots, d$; $\Delta P_a|_M$ – среднее значение эффективной полосы частот воздействия на квазистационарном длительном интервале наблюдения, определяемое согласно следующего выражения:

$$\Delta P_a|_M = \frac{1}{M} \cdot \sum_{m=1}^{m=M} \Delta P(m). \quad (5.24)$$

где M – количество выделенных сегментов паузы РС.

$$\alpha(d) \leq \alpha_{nbn}^o, \quad (5.25)$$

где $\alpha_{nbn}^o = 2$ – пороговое значение стабильности эффективной полосы частот узкополосных акустических воздействий.

Значение α_{nbn}^o определяется исходя из анализа (таблицы 4.1) в которой показано, что для всех узкополосных акустических воздействий оценка отношения максимального значения эффективной полосы частот к среднему не превышает 2.

Невыполнение условия (5.25) } выполнение итеративного алгоритма «с выкалыванием» максимального значения:

$$\begin{aligned} \Delta P^B &= \max[\Delta P]_M, \\ M &= M - 1, \\ \Delta P|_M &= \Delta P|_M - \Delta P^B, \end{aligned} \quad (5.26)$$

где ΔP^B – максимальное значение эффективной полосы частот для всех сегментов, определенных как пауза.

{Выполнение условия (5.25)} оценка порогового значения эффективной полосы частот акустического воздействия ΔP^{tr} :

$$\Delta P^{tr} = \max[\Delta P]_M. \quad (5.27)$$

19-20) Оценка бинарной функции признака усреднения спектральных характеристик акустического воздействия $J_P(m)$ для сегмента d при

$m = d - D_m + 1, \dots, d :$

$$\begin{aligned} J(m) = 0 &\Rightarrow J_P(m) = 0, \\ J(m) = 1 \ \&\ \Delta P(m) \leq \Delta P^{tr} &\Rightarrow J_P(m) = 1, \\ J(m) = 1 \ \&\ \Delta P(m) > \Delta P^{tr} &\Rightarrow J_P(m) = 0. \end{aligned} \quad (5.28)$$

21) Усредненная оценка спектральных характеристик акустического воздействия $\langle A_E(k, d) \rangle$ для сегмента d при $m = d - D_m + 1, \dots, d :$

$$A_U^a(k, m) = A_U(k, m) \cdot J_P(m). \quad (5.29)$$

$$\langle A_E(k, d) \rangle = \frac{1}{\sum_{m=d-D_m+1}^{m=d} J_P(d)} \cdot \sum_{m=d-D_m+1}^{m=d} A_U^a(k, m). \quad (5.30)$$

22) Оценка коэффициента усреднения $I(d)$ для сегмента d и проведение шумоподавления прямым вычитанием в спектрах амплитуд Фурье:

$$J(d) = 0 \Rightarrow I(d) = \frac{1}{2}, \quad (5.31)$$

$$J(d) = 1 \Rightarrow I(d) = 1,$$

$$A_{SS}(k, d) = I(d) \cdot |A_U(k, d) - \langle A_E(k, d) \rangle|, \quad (5.32)$$

$$\dot{A}_{SS}(k, d) = A_{SS}(k, d) \cdot \exp[F_U(k, d)], \quad (5.33)$$

где $\dot{A}_{SS}(k, d)$ – оценка комплексного спектра сигнала на выходе системы фильтрации; $F_U(k, d)$ – оценка спектра фаз сигнала на входе системы.

После проведения шумоподавления получение отсчетов РС на выходе системы фильтрации путем ОДПФ.

23) {Выполнение условия (5.20)} сохранение пороговых значений для КСБ с учетом их определенности на предыдущем сегменте:

$$\lambda_j^{tr}(d) = \lambda_j^{tr}(d-1) \quad (5.34)$$

24-25) Аналогично пунктам 16-17 настоящего алгоритма.

В случае, когда пороговые значения $\lambda_j^{tr}(d-1)$ не определены:

$$\begin{aligned}\lambda_j^{tr}(d) &= [\quad], \\ J(m) &= 0.\end{aligned}\tag{5.35}$$

26) Сохранение оценки спектральных характеристик акустического воздействия:

$$\langle A_E(k, d) \rangle = \langle A_E(k, d-1) \rangle.\tag{5.36}$$

27) Аналогично пункту 22 настоящего алгоритма.

28) {Невыполнение условия (5.5)} оценка КСБ $\lambda_j(d)$ и количества знакоперемен при $\beta_j(d)|_{M=D_m}$ аналогично пункту 11.

29) Проверка выполнения условия по пороговому значению для широкополосного акустического воздействия j -му КСБ:

$$\beta_j(d)|_{M=D_m} \geq \beta_{wbn}^o\tag{5.37}$$

где β_{wbn}^o – пороговое значение количества знакоперемен для широкополосного акустического воздействия.

30-34) Аналогично пунктам 13-17 настоящего алгоритма.

35) Аналогично пункту 18 настоящего алгоритма.

36-37) Оценка бинарной функции признака усреднения спектральных характеристик широкополосного акустического воздействия при $m = d - D_m + 1, \dots, d$:

$$\begin{aligned}J(m) = 0 &\Rightarrow J_P(m) = 0, \\ J(m) = 1 \ \&\ \ 0,9 \cdot \frac{F_d}{2} - \xi \leq \Delta P(m) \leq 0,9 \cdot \frac{F_d}{2} + \xi &\Rightarrow J_P(m) = 1, \\ J(m) = 1 \ \&\ \ 0,9 \cdot \frac{F_d}{2} - \xi > \Delta P(m) > 0,9 \cdot \frac{F_d}{2} + \xi &\Rightarrow J_P(m) = 0.\end{aligned}\tag{5.38}$$

38) Аналогично пункту 21 настоящего алгоритма.

39) Идеализация усредненной оценки спектральных характеристик широкополосного акустического воздействия $\langle A_E^{id}(k, d) \rangle$:

$$\langle A_E^{id}(k, d) \rangle = \frac{1}{N-1} \cdot \sum_{k=0}^{k=N-1} \langle A_E(k, d) \rangle.\tag{5.39}$$

40-45) Аналогично пунктам 22-27 настоящего алгоритма.

Результативность представленного метода показана на (рисунке 5.2), а его

сравнение с известными техническими решениями отражено в разделе 5.3. В качестве тестового материала использовались фразы рекомендованные [84].

Анализ блок схемы алгоритма адаптивной цифровой фильтрации с фиксированным сегментом анализа (А.1) показывает, что данный подход из-за постоянной длительности обрабатываемой последовательности не имеет совокупности усредненных оценок спектральных характеристик акустического воздействия, полученных при различных значениях спектрального разрешения. Решение этого вопроса лежит в плоскости переменного спектрального разрешения и адаптации к параметрическим особенностям речевого сигнала.

В теоретическом плане имеется возможность применения дискретного преобразования Фурье с неравномерным спектральным разрешением [39, 145, 146]. Однако, данный подход требует сложного механизма адаптации по многочисленным параметрам РС с целью выбора оптимального спектрального разрешения, кроме того грани адаптации размываются в условиях интенсивного зашумления, т.е. спектральное разрешение в значительной степени будет зависеть от параметров акустического воздействия, а не от характеристик речи. С целью решения вопросов реализации переменного спектрального разрешения в ходе исследований в рамках представленного метода был разработан алгоритм адаптивной цифровой фильтрации РС с переменным сегментом анализа (далее А.2) [141], который позволяет добиться конечной по мощности совокупности спектральных разрешений, увеличение начального значения которых достигается делением обрабатываемого сегмента (рисунок 5.3).

Разработанный алгоритм является случаем разветвленной обработки речевого сигнала с различными размерами сегментов локальной стационарности, каждая ветвь которого, в свою очередь, является независимым алгоритмом адаптивной цифровой фильтрации с фиксированным сегментом анализа.

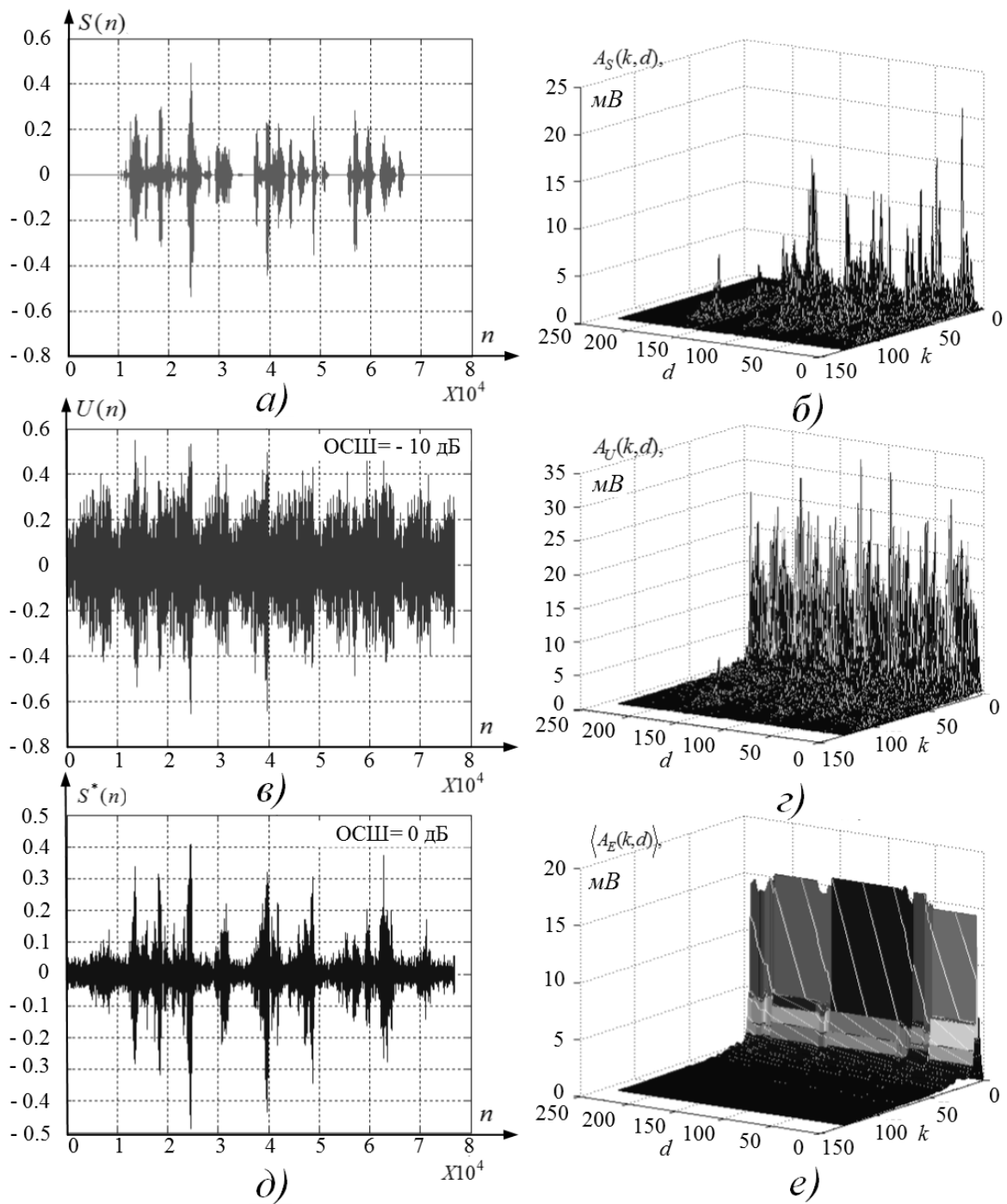


Рисунок 5.2 – Результативность алгоритма адаптивной фильтрации РС с фиксированным сегментом анализа (шум вертолета ОСШ = минус 10 дБ):

- а) фраза длительностью 10 секунд в отсутствии зашумления; б) оценка спектра амплитуд в отсутствии зашумления посегментно; в) РС на входе системы фильтрации (среднее ОСШ в течении 10 секунд – 10 дБ); г) оценка спектра амплитуд РС посегментно в УИАВ; д) сигнал на выходе системы фильтрации (среднее ОСШ в течении 10 секунд 0 дБ); е) усредненная оценка спектра амплитуд акустического воздействия посегментно.

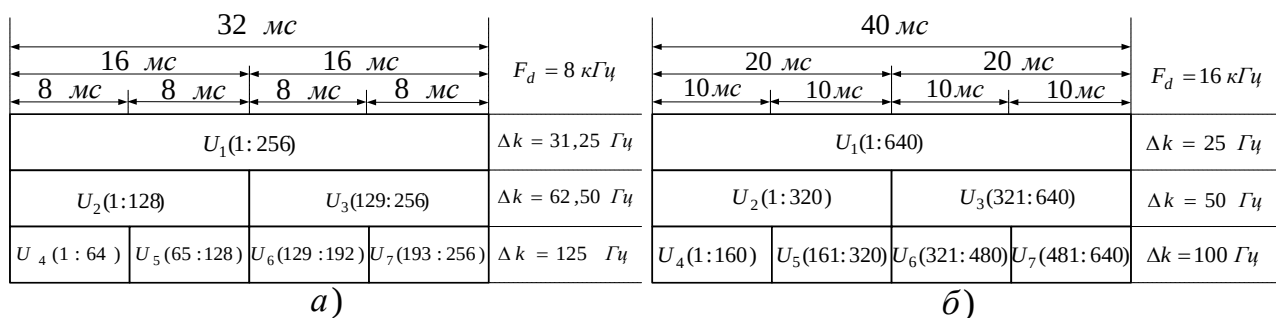


Рис. 5.3 – Процедура деления сегмента анализа

Анализ этапов выполнения алгоритма, представленного на (рисунке 5.4) можно описать следующим образом:

I) Ввод отсчетов $U(1:N,d)$ дискретного сигнала обрабатываемого сегмента квазистационарности d , где N – количество отсчетов на сегменте локальной стационарности. Сброс счетчиков количества итераций по каждой у-независимой ветви обработки РС.

II) Деление обрабатываемого сегмента локальной стационарности (рисунок 5.3), выполняемое в блоке 2 (рисунок 5.4).

III) Выполнение алгоритма адаптивной цифровой фильтрации РС с фиксированным спектральным разрешением (А.1) для каждой у-независимой ветви обработки:

- блоки 3, 6, 9, 12, 15, 18, 21 оценка пороговых значений по каждому КСБ для у-независимой ветви обработки, оценка бинарной функции признака паузы;

- блоки 4, 7, 10, 13, 16, 19, 22 стабилизация и усредненная оценка спектральных характеристик акустического воздействия, оценка коэффициента усреднения для каждой из у-независимых ветвей обработки;

- блоки 5, 8, 11, 14, 17, 20, 23 выполнение шумоподавления с применением вычитания в спектрах амплитуд Фурье при учете психоакустического восприятия зашумленной речи слуховым аппаратом человека, получение отсчетов РС на выходе для каждой у-независимой ветви обработки путем ОБПФ.

IV) Получение значений отсчетов РС на выходе системы фильтрации путем усреднения результатов у-независимых ветвей обработки сегмента d РС согласно уровневой структуры деления (рисунок 5.3), выполняемое в блоке 25 рисунка 5.4.

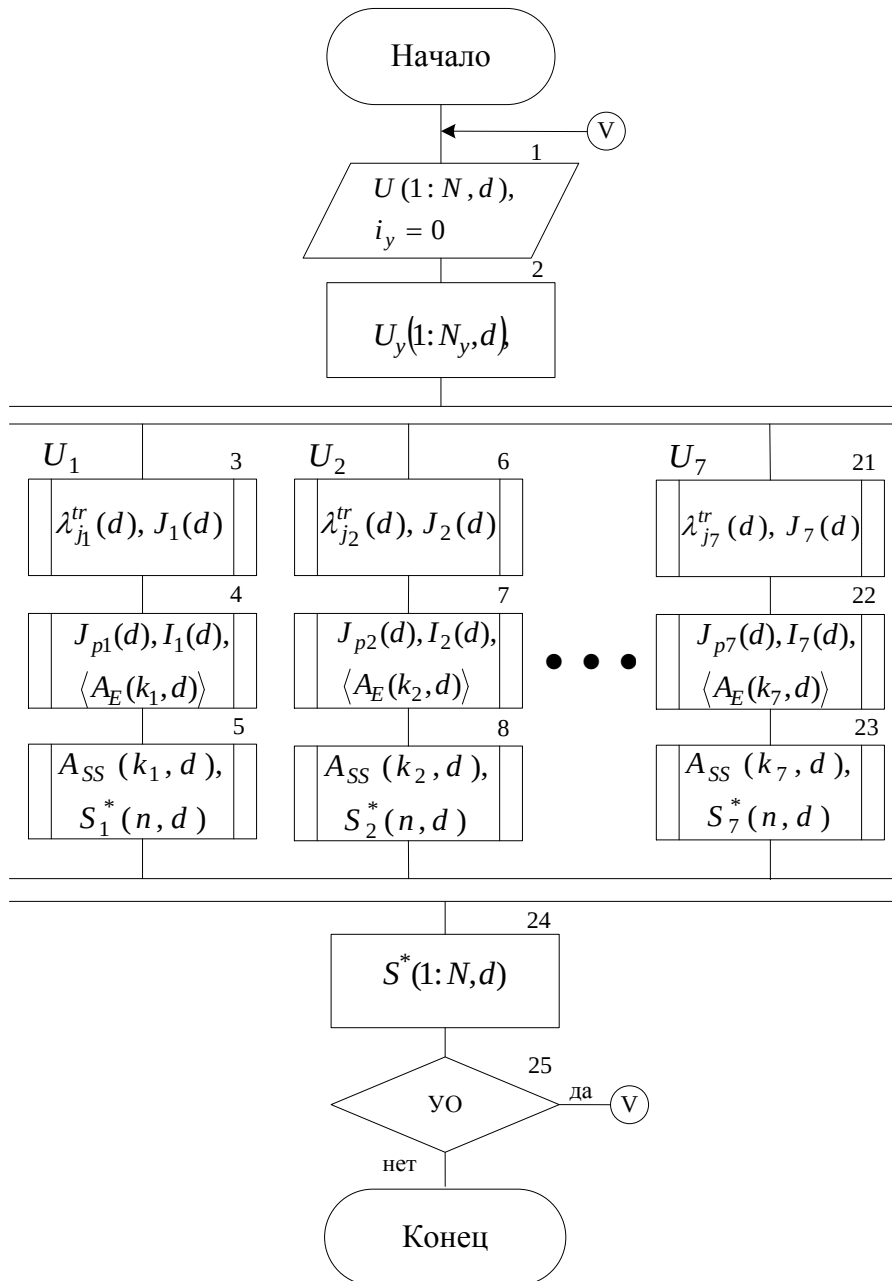


Рис. 5.4 – Блок-схема алгоритма адаптивной цифровой фильтрации речевого сигнала с переменным сегментом анализа

Необходимо отметить, что при сравнительном анализе А.1 и А.2 в ходе исследований выявлено улучшение субъективной оценки качества РС на выходе системы фильтрации для А.2.

Анализ полученных посредством функционирования представленного алго-

ритма шумоподавления результатов позволяет говорить о том, что введение в системы анализа РС процедур предобработки значительно повышает значение ОСШ, восстанавливает значения ЛСЧ и приближает статистические свойства очищенного сигнала к значениям РС до зашумления. С учетом вышеизложенного можно сделать вывод о целесообразности применения разработанных алгоритмов адаптивной цифровой фильтрации, а также методе шумоподавления, представленного в разделе 4.2 в качестве процедуры предобработки на начальном этапе различных методов анализа и обработки РС.

5.3 Критерий оценки качества речевого сигнала после его обработки

В настоящее время при оценивании качества передачи РС применяются различные методы и критерии, но не один из них не является универсальным [188, 235]. Проанализировав многообразие методов и критериев оценивания качества передачи РС, можно сделать следующие выводы:

- невозможность сведения результатов измерения физических параметров исследуемого тракта к единому количественному показателю обуславливает отказ от использования физических методов;
- превалирование интересов абонентов среди факторов, определяющих тенденции развития систем обработки РС в системах инфокоммуникаций, делают более актуальными методы субъективной оценки, характерной особенностью которых является участие в них бригад дикторов и аудиторов с определением последними качества звучания речи, при этом качество речи определяется величиной, характеризующей субъективную оценку звучания речи на выходе УОРС;
- субъективные тесты являются дорогостоящими, время- и трудоемкими, исключают возможность автоматизации процесса измерения и требуют специальной подготовки квалифицированной группы экспертов, поэтому исследователи непрерывно ведут поиск объективных критериев и методов оценивания качества

звучания восстановленного РС, хорошо коррелирующий с результатами субъективных тестов;

- состоятельность показателя оценки качества речи в значительной степени зависит от выбора критерия оценки;
- внедрение новых технологий обработки речи и возросшие требования абонентов к качеству передачи РС привели к расширению перечня показателей качества речи: наравне с традиционными (громкость и разборчивость) все более активное применение находят естественность звучания голоса и узнаваемость говорящего;
- для оценивания качества передачи РС необходимы методы, позволяющие получить обобщенную интегральную оценку качества речи с учетом ее громкости, разборчивости, естественности звучания и узнаваемости;
- объективные методы с использованием психо-акустической модели слуховой системы человека позволяют получить устойчивые оценки качества передачи РС, но характеризуются сложностью реализации перцептуальной модели слухового аппарата человека;
- корректность показателей объективной оценки качества передачи РС характеризуется корреляцией оценок, вычисленных с их использованием, с оценками, полученными субъективными методами.

Таким образом, работоспособность метода и критерия объективного оценивания качества РС на входе и выходе УОРС, ориентированных на восприятие аудиторной системой человека будет определяться их способностью быстро и надежно предоставлять оценку, максимально предсказывающую величину общепринятой субъективной оценки качества. Стандартом оценки качества передачи РС является оценка группой слушателей, получаемая прослушиванием в субъективных тестах, которые проводятся в соответствии с рекомендациями сектора стандартизации в области телефонии Международного союза электросвязи [66]. Результаты этих тестов – усредненные значения средней экспертной оценки. Шкала средней субъективной оценки MOS, определяющая пять классов РС приведена в (таблице 5.1). Качество РС оценивают путем сравнения с эталонным сигналом, в качестве которого

используют стандартный телефонный тракт. Оценку выполняют по контрольным фразам, приведенным в приложении А [84]. Бригада auditors должна состоять из 10—12 человек. Аудиторами не могут быть специалисты по КРС. Число дикторов должно быть не менее 5 (мужчины и женщины), при этом каждую контрольную фразу диктор произносит два раза — один раз через оцениваемый тракт, другой — через эталонный тракт. Методы измерений и нормы качества передачи речи регламентированы в [83, 84]. В данной рекомендации представлен метод парных сравнений для оценивания качества обработки РС, (таблица 5.2). Объективные методы используют некоторые формализованные параметры, позволяющие судить о степени различия между РС на входе и выходе УОРС. Объективные измерения эффективны с точки зрения цены и повторяемости результатов, позволяют исследователю получить быструю и надежную оценку качества речи непосредственно в процессе разработки системы фильтрации РС. Оценка меры зашумления РС при этом может проводиться как во временной, так и в частотной областях [184]. Наиболее известными измерениями степени зашумления во временной области являются СКО, показатель ОСШ.

Таблица 5.1 – Оценки Mean Opinion Score

Оценка	Качество	Искажения
5	Превосходно	Незначительные
4	Хорошо	Заметны, но не раздражающие
3	Посредственно	Немного раздражают
2	Плохо	Раздражают, но не неудобны
1	Неудовлетворительно	Очень раздражают

В рамках диссертационных исследований объективным критерием качества на первом этапе оценки РС на выходе УОРС (вход кодера, выход декодера), функционирующего по тому или иному алгоритму согласно целевой функции разработанной модели (2.63) выбрано среднее повышение ОСШ выход-вход. Однако исследования, проведенные в рамках современных требований абонентов по качеству речевого сигнала, показывают, что ОСШ лишь раскрывает меру зашумления РС, но не отражает в полной мере субъективную оценку качества. Для устранения вышеназванных недостатков, возникающих при оценивании качества передачи РС,

применяют критерии в частотной области, такие как сегментно-спектральный критерий (*Segmental Spectral Distortion* (SSD)), расстояние Итакуры-Саито ISD (*Itakura-Saito Distance*), кепстральное расстояние CD (*Cepstral Distance*). Главным недостатком указанных методов является зависимость значений коэффициентов корреляции между оценками ISD, CD и MOS от типа и скорости передачи оцениваемой системы.

Таблица 5.2 – Критерии оценки для метода парных сравнений

Характеристика качества речи	Баллы
Естественность звучания речи. Высокая узнаваемость. Полное отсутствие помех и искажений	4.6-5.0
Естественность звучания речи. Высокая узнаваемость. Отдельные малозаметные искажения или помехи	4.0-4.5
Естественность звучания речи. Высокая узнаваемость. Слабое постоянное присутствие отдельных видов искажений или помех	3.5-3.9
Незначительное нарушение естественности и узнаваемости. Заметное присутствие отдельных искажений или помех	3.0-3.4
Заметное нарушение естественности и ухудшение узнаваемости, присутствие нескольких видов искажений (картавость, гнусавость и др.) или помех	2.5-2.9
Существенное искажение естественности и ухудшение узнаваемости. Постоянное присутствие искажений или помех	1.7-2.4
Сильные искажения типа картавость, гнусавость или наличие сильных помех. Наблюдается потеря естественности и узнаваемости	< 1.7

Перспективными являются психо-акустические методы и критерии оценивания качества передачи РС, основанные на моделях слуха человека. Данные методы направлены на имитацию с помощью технических и программных средств алгоритма слуховой перцепции человека. Это объективные измерения, но уже с учетом особенностей восприятия речи человеком.

Наиболее распространенными критериями является искажение спектра барков BSD (*Bark Spectral Distortion*), модифицированное искажение спектра барков MBSD, алгоритм PSQM (*Perceptual Speech Quality Measure*) и алгоритм PESQ (*Perceptual Evaluation of Speech Quality*). Необходимо отметить, что объективные методы, использующие психо-акустическую модель слуховой перцепции аудиторной системы человека, обеспечивают высокие значения коэффициентов корреляции с MOS.

Для обоснования критерия качества звучания очищенного речевого сигнала на втором

этапе были получены оценки критерия качества речевого сигнала на выходе системы обработки, включающие следующие пункты:

1) Получение средних субъективных оценок качества зашумленного и очищенного РС для различных уровней зашумления и различных видов шумов.

2) Получение средних объективных оценок качества звучания зашумленного и очищенного РС для различных отношений сигнал-шум на входе. Оценка мер искажений РС при этом проводилась как во временной, так и в частотной областях. Вычислялись следующие меры искажений: ОСШ, SSD, ISD, CD и MBSD.

3) Определение коэффициентов корреляции между объективными и субъективными оценками для выбранных мер искажений (таблица 5.3) согласно следующему выражению:

$$r = \frac{\sum_{h=1}^W \sum_{i=1}^Z (X^j(i) - \bar{X})(Y^j(i) - \bar{Y})}{\sqrt{\sum_{h=1}^W \sum_{i=1}^Z (X^j(i) - \bar{X})^2 \cdot \sum_{i=1}^Z (Y^j(i) - \bar{Y})^2}}, \quad (5.40)$$

где $X(i)$ и $Y(i)$ – объективный и субъективный показатели качества очищенного и зашумленного РС при j -м шуме; $\bar{X}$ и $\bar{Y}$ – средние значения объективного и субъективного показателей качества на множестве объектов исследования; W – число исследуемых шумов, Z – число уровней ОСШ.

Таблица 5.3 – Оценка коэффициента корреляции

Методы оценивания качества	Коэффициент корреляции r
Критерий отношения сигнал/шум, ОСШ	0,74
Сегментно-спектральный критерий, SSD	0,68
Расстояние Итакуры-Саито, ISD	0,61
Кепстральное расстояние, CD	0,63
Модифицированное искажение спектра барков, MBSD	0,87

Учитывая полученные результаты (таблица 5.3) расчета коэффициента корреляции между объективными и субъективными оценками, можно выделить показатель MBSD ($r = 0,87$) как более точную меру субъективного качества РС. Указанный показатель определяется как:

$$\text{MBSD} = \frac{1}{D^{\max}} \sum_{d=1}^{D^{\max}} \sum_{j=1}^{J^{\max}} M(d, j) \cdot D_{xy}(d, j), \quad (5.41)$$

где $M(d, j)$ и $D_{xy}(d, j)$ – j -е параметры искажений уровня ощущения и разницы уровней громкости d -го сегмента анализа; $D^{\max}$ – число сегментов в речевом фрагменте; $J^{\max}$ – общее количество критических полос [237].

Коэффициент корреляции между субъективными оценками и объективной оценкой через анализ ОСШ оказался равным $r = 0,74$ однако данный показатель имеет недооценку субъективного восприятия речи в диапазоне ОСШ от 15 до 0дБ. Поэтому при оценке повышения качества обрабатываемого РС предложенных алгоритмических решений выбрано два критерия – это среднее повышение ОСШ на выходе системы фильтрации по сравнению с известными решениями и среднее снижение показателя MBSD [205, 237], как объективную меру субъективного восприятия речи (таблица 5.4).

Таблица 5.4 – Оценки качества речевого сигнала MOS и MBSD

Характеристика качества речи	Q^{out} Баллы	MBSD
Естественность звучания речи. Высокая узнаваемость. Полное отсутствие помех и искажений	4.6-5.0	$\leq 0,3$
Естественность звучания речи. Высокая узнаваемость. Отдельные малозаметные искажения или помехи	4.0-4.5	$> 0,3 \leq 1,5$
Естественность звучания речи. Высокая узнаваемость. Слабое постоянное присутствие отдельных помех	3.5-3.9	$> 1,5 \leq 3,2$
Незначительное нарушение естественности и узнаваемости. Присутствие слабых помех	3.0-3.4 Q^{tresh}	$> 3,2 \leq 5,7$
Заметное нарушение естественности и ухудшение узнаваемости, присутствие искажений или помех	2.5-2.9	$> 5,7 \leq 9,3$
Существенное искажение естественности и ухудшение узнаваемости. Постоянное присутствие помех	1.7-2.4	$> 9,3 \leq 12,7$
Сильные искажения типа картавость, гнусавость или наличие сильных помех. Механический голос. Наблюдается потеря естественности и узнаваемости	< 1.7	$> 12,7$

5.4 Оценка качества речевого сигнала при подавлении помех согласно предлагаемым техническим решениям

Необходимо отметить, что при допущении, выражаемом идеализацией функционирования системы обработки речевой информации от входа кодера до выхода декодера с коэффициентом передачи, равным единице (при учете составного соединения системы фильтрации и системы кодирования в рамках УОРС) оценка качества РС на выходе системы фильтрации имеет равнозначный характер с оценкой на выходе декодера. При этом сравнительный анализ того или иного алгоритма необходимо выполнять без учета оценки качества РС в существующих УОРС, т.е. без системы фильтрации.

При проведении фильтрации в условиях незначительного зашумления РС известные технические решения имеют высокую эффективность, однако в условиях интенсивного зашумления наблюдается ее снижение, вследствие появления ситуации, связанной с трудностью оценки спектральных характеристик акустического воздействия и возникновением нелинейных искажений после проведения спектрального вычитания, что снижает в итоге ОСШ и значительно ухудшает субъективное восприятие речи.

Исходя из этого объективным критерием качества очищенной речи на выходе системы фильтрации, функционирующей согласно предложенным алгоритмам, относительно известных решений на первом этапе выбрано среднее повышение ОСШ в диапазонах от 35 до -5 дБ (диапазон робастности разработанных алгоритмических решений), от 15 до -5 дБ (диапазон условий интенсивных акустических воздействий), от 35 до 15 дБ (диапазон условий отсутствия интенсивного зашумления).

Также приведен коэффициент относительного ОСШ G_j^{opt} , % (при сравнении с оптимумом в рамках предложенной математической модели) определяемого согласно следующему выражению:

$$G_j^{opt} = \sum_{h=1}^{h=W} \sum_{i=1}^{i=Z} \frac{\Delta G_{out}^j(h, i)}{\Delta G_{out}^{opt}(h, i)} \cdot 100\%, \quad (5.42)$$

где $\Delta G_{out}^{opt}(h, i)$ – среднее максимальное значение разницы ОСШ на выходе системы фильтрации для всех дикторов и всех записей от i -го значения ОСШ на входе и h -м виде шума (в рамках предложенной модели при равенстве усредненной оценки и истинной оценки спектра амплитуд акустического воздействия $\langle A_E(k, d) \rangle = A_E(k, d) \quad \forall d$); Z – количество уровней ОСШ в выбранном диапазоне; W – количество исследуемых шумов; $\Delta G_{out}^j(h, i)$ – значение разницы ОСШ на выходе системы фильтрации, функционирующей по j -му алгоритму для всех дикторов и всех записей от i -го значения ОСШ на входе и h -м виде шума, определяемая как:

$$\Delta G_{out}^j(h, i)[\partial B] = G_{out}^j(h, i) - G_{in}^j(h, i), \quad (5.43)$$

где $G_{in}^j(h, i)$ – i -е значение ОСШ на входе системы фильтрации для всех дикторов и всех записей h -го вида шума.

Среднее повышение качества обрабатываемого сигнала измеряется как в процентном соотношении, так и в абсолютных значениях дБ согласно следующим выражениям:

$$\Delta G_{A.1 \rightarrow spm}[\partial B] = \frac{1}{W \cdot Z} \sum_{h=1}^{h=W} \sum_{i=1}^{i=Z} (\Delta G_{out}^{A.1}(h, i) - \Delta G_{out}^{spm}(h, i)), \quad (5.44)$$

$$\Delta G_{A.2 \rightarrow spm}[\partial B] = \frac{1}{W \cdot Z} \sum_{h=1}^{h=W} \sum_{i=1}^{i=Z} (\Delta G_{out}^{A.2}(h, i) - \Delta G_{out}^{spm}(h, i)), \quad (5.45)$$

$$\Delta G_{A.2 \rightarrow A.1}[\partial B] = \frac{1}{W \cdot Z} \sum_{h=1}^{h=W} \sum_{i=1}^{i=Z} (\Delta G_{out}^{A.2}(h, i) - \Delta G_{out}^{A.1}(h, i)), \quad (5.46)$$

$$\Delta G_{A.1 \rightarrow spm}^{opt} [\%] = G_{A.1}^{opt} - G_{spm}^{opt}, \quad (5.47)$$

$$\Delta G_{A.2 \rightarrow spm}^{opt} [\%] = G_{A.2}^{opt} - G_{spm}^{opt}, \quad (5.48)$$

$$\Delta G_{A.2 \rightarrow A.1}^{opt} [\%] = G_{A.2}^{opt} - G_{A.1}^{opt}, \quad (5.49)$$

где $G_{out}^j(h, i)$ – значение ОСШ на выходе системы фильтрации, функционирующей по j -алгоритму при i -м значении ОСШ на входе и h -м виде шума (А.1 – разработанный алгоритм с фиксированным сегментом анализа, А.2 – разработанный алгоритм с переменным сегментом анализа, spm – алгоритм фильтрации с использованием вычитания в спектрах мощности, выбранный для совершенствования); Z – количество уровней ОСШ на входе системы фильтрации в выбранном диапазоне; W – количество видов акустических воздействий: (1. квазибелый гауссовский шум; 2. шум двигателя; 3. шум города; 4. шум ветра; 5. шум вертолета; 6. тоновой шум (суперпозиция трех синусоид 450, 950, 1100 Гц); 7. шум проезжающего поезда; 8. шум боя; 9. шум горящего здания; 10. шумовая смесь (шум двигателя, шум города, шум ветра))

На втором этапе использовался показатель MBSD [237], как наиболее коррелированный с субъективной оценкой качества РС. Оценку среднего снижения показателя MBSD ΔX_j на выходе системы фильтрации, функционирующей по j -алгоритму будем проводить согласно следующим выражениям:

$$\Delta X_{A.1 \rightarrow spm} = \frac{1}{W \cdot Z} \sum_{h=1}^{h=W} \sum_{i=1}^{i=Z} (X_{out}^{spm}(h, i) - X_{out}^{A.1}(h, i)), \quad (5.50)$$

$$\Delta X_{A.2 \rightarrow spm} = \frac{1}{W \cdot Z} \sum_{h=1}^{h=W} \sum_{i=1}^{i=Z} (X_{out}^{spm}(h, i) - X_{out}^{A.2}(h, i)), \quad (5.51)$$

$$\Delta X_{A.2 \rightarrow A.1} = \frac{1}{W \cdot Z} \sum_{h=1}^{h=W} \sum_{i=1}^{i=Z} (X_{out}^{A.1}(h, i) - X_{out}^{A.2}(h, i)), \quad (5.52)$$

где $X_{out}^j(h, i)$ – значение оценки MBSD на выходе системы фильтрации, функционирующей по j -алгоритму i -м значении ОСШ на входе и h -м виде шума (А.1 – разработанный алгоритм с фиксированным сегментом анализа, А.2 – разработанный алгоритм с переменным сегментом анализа, *spt* – алгоритм фильтрации с использованием вычитания в спектрах мощности, выбранный для совершенствования); Z – количество уровней ОСШ на входе системы фильтрации в выбранном диапазоне; W – количество видов акустических воздействий.

Оценка алгоритмов проводилась на зашумленных записях, особенности формирования которых представлены в Приложении 1.

В ходе экспериментальной проверки повышения качества обрабатываемого РС, получаемого на выходе системы фильтрации, функционирующей по разработанным алгоритмам относительно известного и выбранного для совершенствования алгоритма фильтрации с использованием вычитания в спектрах мощности, получены следующие результаты (таблица 5.5).

Проведенная оценка повышения качества обрабатываемого РС на входе кодера и анализ (таблицы 5.6) показывают, что разработанные алгоритмы позволяют осуществлять фильтрацию в задаче шумоподавления в условиях интенсивных акустических воздействий со средним повышением ОСШ от 4.00 до 4.34 дБ, средним повышением степени достижения оптимума в рамках предложенной модели от 41,1 до 45, 3 % и средним снижением показателя MBSD от 2.88 до 4,20 (среднее повышение субъективной оценки от 0,87 до 1,22 балла) в диапазоне входных значений ОСШ от 15 до -5 дБ относительно системы фильтрации с применением вычитания в спектрах мощности, взятой для совершенствования.

Таблица 5.5 – Оценка повышения качества обрабатываемого
РС на выходе системы фильтрации

Вид шума		Динамический диапазон ОСШ								
		от 35 до 15 дБ			от 15 до – 5 дБ			от 35 до – 5 дБ		
		G^{opt} %	ΔG_{out} дБ	X_{out} мсон	G^{opt} %	ΔG_{out} дБ	X_{out} мсон	G^{opt} %	ΔG_{out} дБ	X_{out} мсон
КБГШ	A.1	86	1,78	1,30	67	5,46	4,27	66	3,90	2,87
	A.2	99	2,07	1,10	74	6,05	2,87	75	4,40	1,88
	spm	61	1,05	2,14	38	2,4	6,84	39	2,12	4,70
Шум двигателя	A.1	82	1,92	1,62	65	6,00	4,14	64	4,28	2,65
	A.2	93	2,15	0,95	70	6,46	2,45	70	4,67	1,65
	spm	50	1,13	2,23	21	1,82	6,96	23	1,47	4,45
Шум города	A.1	52	2,69	1,42	38	4,86	4,38	40	3,79	2,77
	A.2	67	2,97	0,98	45	5,02	2,98	42	3,88	1,87
	spm	19	0,97	2,04	10	1,38	7,23	12	1,15	4,43
Шум ветра	A.1	81	2,01	1,25	58	5,14	3,37	59	3,77	2,98
	A.2	86	2,13	0,99	61	5,46	2,56	63	4,05	1,58
	spm	34	0,86	2,06	15	1,33	6,54	17	1,08	4,89
Шум горящего здания	A.1	65	3,00	1,39	49	5,88	4,03	50	4,51	2,65
	A.2	66	3,05	1,13	51	6,07	2,65	52	4,67	1,69
	spm	24	1,14	2,21	13	1,56	7,12	14	1,32	4,93
Шум проезжающего поезда	A.1	80	1,73	1,35	57	5,06	4,11	57	3,60	2,66
	A.2	91	1,95	1,17	61	5,41	2,45	62	3,91	1,87
	spm	34	0,74	1,87	13	1,20	6,72	15	0,97	4,98
Шум вертолета	A.1	86	2,45	1,43	60	6,08	3,96	60	4,40	2,51
	A.2	92	2,61	1,07	62	6,26	2,85	63	4,43	1,76
	spm	42	1,19	1,65	18	1,87	6,98	20	1,51	4,83
Тоновой шум	A.1	98	2,61	1,33	91	8,36	3,60	92	6,18	2,34
	A.2	99	2,66	0,74	94	8,64	1,54	95	6,32	1,45
	spm	92	2,40	1,46	45	4,17	5,46	49	3,31	4,12
Шумовая смесь	A.1	79	2,76	1,32	60	6,66	3,67	60	4,91	2,57
	A.2	86	3,01	0,97	65	7,30	2,33	65	5,43	1,54
	spm	32	1,11	2,34	14	1,62	6,89	16	1,34	4,58

Результаты применения шумоподавления в РС представлены в виде зависимостей на (рисунки 5.5-5.8).

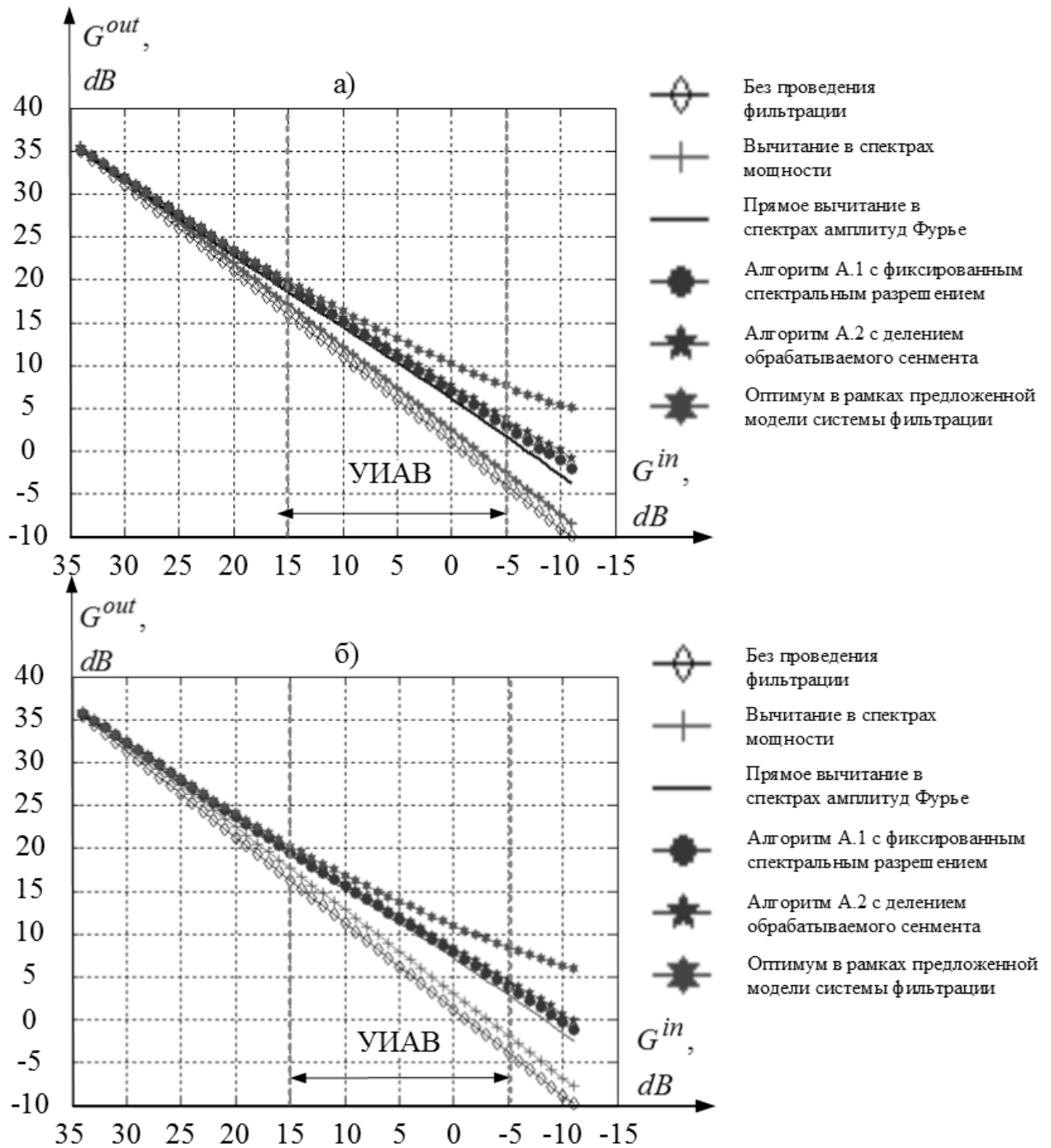


Рисунок 5.5 – Результативность алгоритмов фильтрации речевого сигнала:

- а) зависимость ОСШ на выходе системы фильтрации от ОСШ на входе при выполнении различных алгоритмов для случая зашумления широкополосным акустическим воздействием (на примере шума автомагистрали);
- б) зависимость ОСШ на выходе системы фильтрации от ОСШ на входе при выполнении различных алгоритмов для случая узкополосного воздействия (на примере шума двигателя грузового автомобиля).

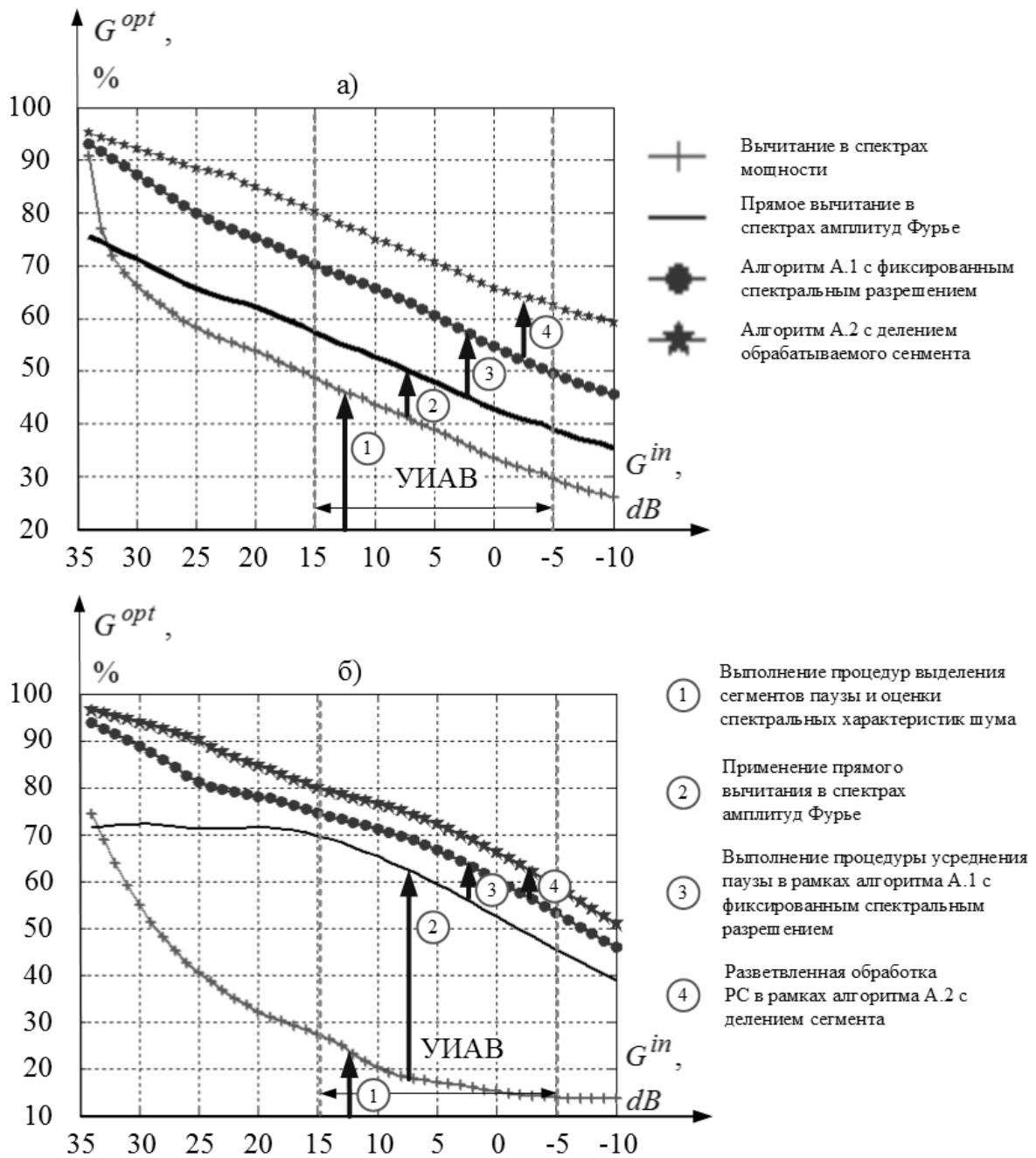


Рисунок 5.6 – Результативность алгоритмов фильтрации речевого сигнала в процентном соотношении достижения оптимума в рамках предложенной модели: а) зависимость ОСШ на выходе системы фильтрации от ОСШ на входе в процентном соотношении относительно оптимума при выполнении различных алгоритмов для случая зашумления широкополосным акустическим воздействием (на примере шума автомагистрали); б) зависимость ОСШ на выходе системы фильтрации от ОСШ на входе при выполнении различных алгоритмов в процентном соотношении относительно оптимума для случая воздействия узкополосного воздействия (на примере шума двигателя).

⊖ Без проведения
фильтрации

+ Вычитание в
спектрах мощностей

● Алгоритм А.1 с фиксированным
спектральным разрешением

★ Алгоритм А.2 с делением
обрабатываемого сегмента

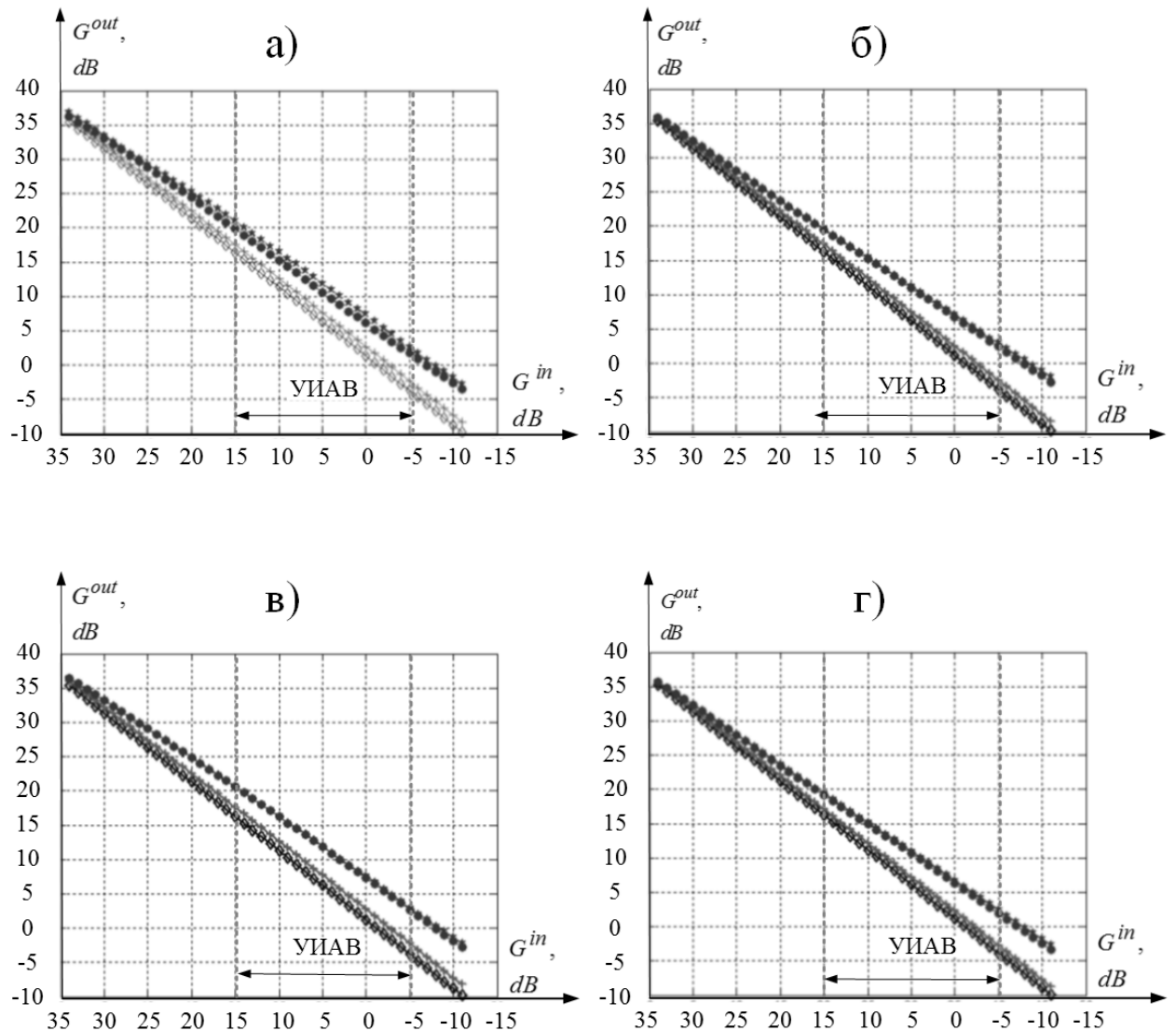


Рисунок 5.7 – Результативность системы фильтрации речевого сигнала при выполнении различных алгоритмов для случая зашумления реальными акустическими воздействиями через зависимость ОСШ на выходе от ОСШ на входе: а) шум города; б) шум ветра; в) шум горящего здания; г) шум боя.

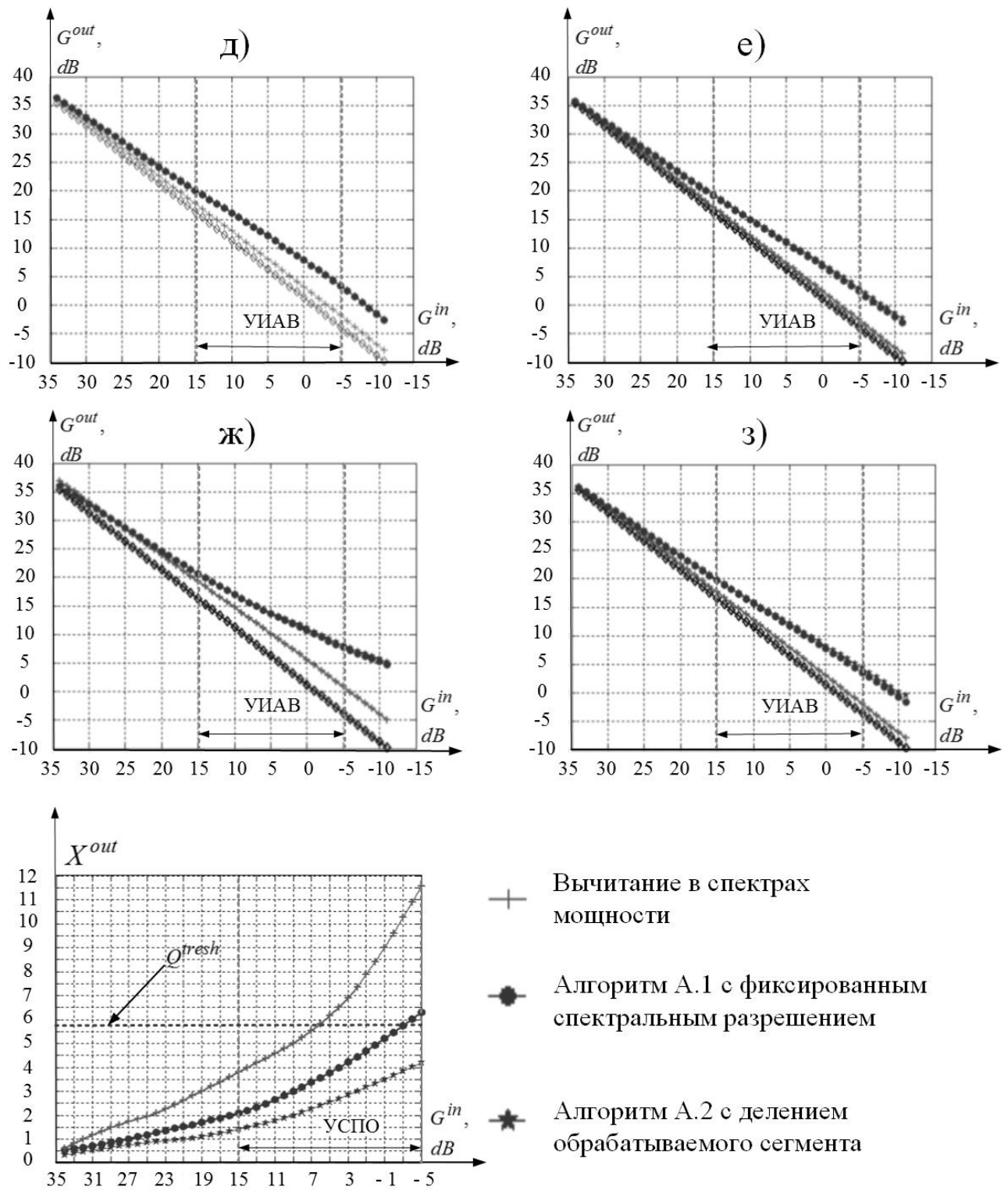


Рисунок 5.8. Результативность системы фильтрации речевого сигнала при выполнении различных алгоритмов для случая зашумления реальными акустическими воздействиями: д) шум проезжающего поезда; е) шум вертолета; ж) тоновой шум; з) шумовая смесь (шум двигателя, города, ветра); и) зависимость усредненной оценки MBSD для всех видов шумов.

Таблица 5.6 – Сравнительная оценка разработанных алгоритмических решений

Сравниваемые алгоритмы	Параметр	Динамический диапазон ОСШ		
		от 35 до 15 дБ	от 15 до - 5 дБ УИАВ	от 35 до - 5 дБ
$A.1 \succ spm$	ΔG , дБ	1,01	4,00	2,78
	ΔG^{opt} , %	36,8	41,1	38,8
	ΔX	0,88	2,88	1,89
$A.2 \succ spm$	ΔG , дБ	1,25	4,34	3,05
	ΔG^{opt} , %	42,6	45,3	43,3
	ΔX	1,27	4,20	2,74
$A.2 \succ A.1$	ΔG , дБ	0,18	0,35	0,26
	ΔG^{opt} , %	5,8	4,3	4,5
	ΔX	0,41	1,32	0,85

Полученные результаты указывают на повышение точности выполнения оценки значений параметров элементов декомпозиции РС в процессе его обработки на основе метода ЛП с ВК его параметров при применении данных технических решений по шумоподавлению в РС. Правильная идентификация векторов КК параметров передаточной функции голосового тракта и соответствующего сигнала возбуждения дополнительно приведет к подавлению остаточного акустического шумового воздействия, так как значения данных векторов обучены на фрагментах активного РС при отсутствии побочных шумовых составляющих различной природы. Положительный эффект от предобработки РС с целью шумоподавления проиллюстрирован на (рисунках 5.9–5.16). Анализ (рисунков 5.9–5.16) показывает, что после проведения шумоподавления в очищенном сигнале (как для случая узкополосного, так и для широкополосного шумового воздействия) происходит восстановление состава локальных мод огибающих спектра амплитуд (ЛМОСА), что приведет на сегменте анализа к более точному выделению линейных спектральных частот РС.

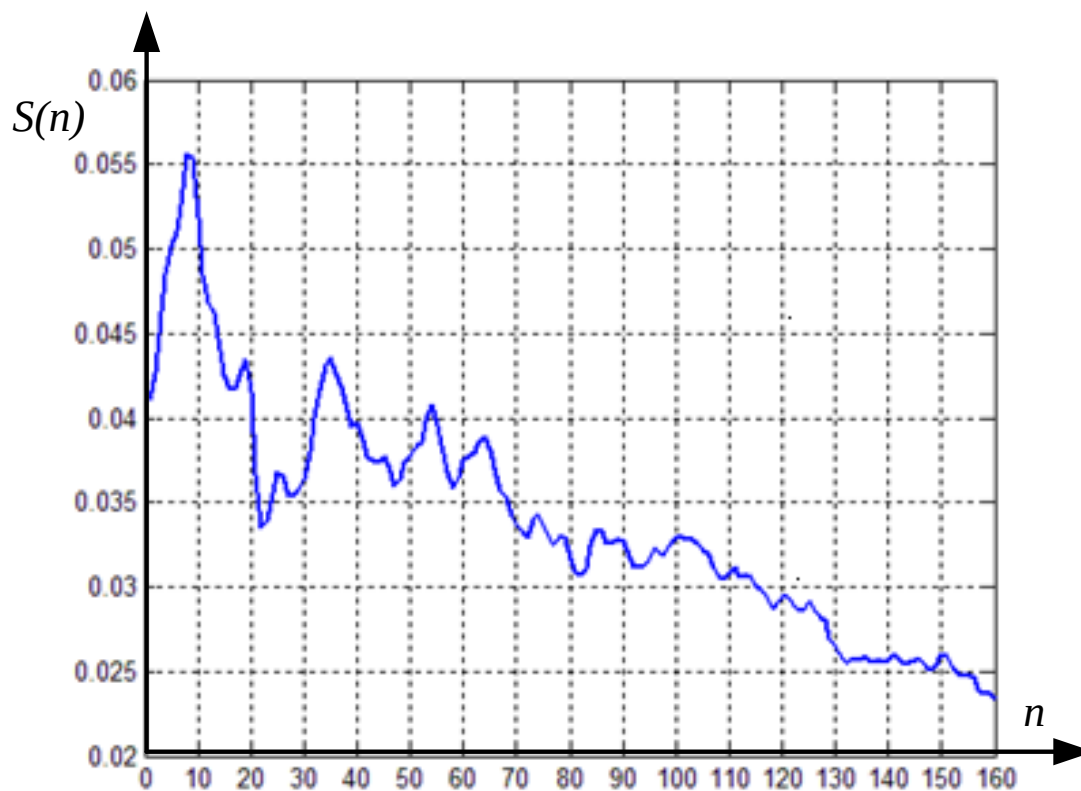


Рисунок 5.9 – Чистый сегмент РС, описывающий переходный элемент речи с низкой энергетикой

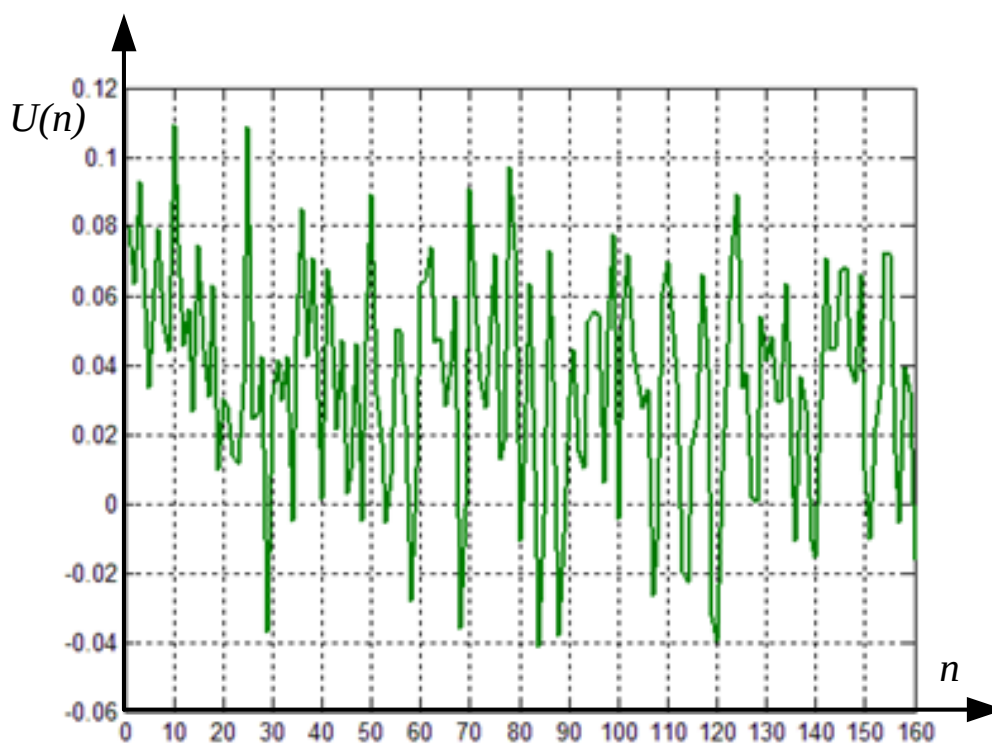


Рисунок 5.10 – Зашумленный БГШ сегмент РС ОСШ = 0 дБ (измерение ОСШ осуществляется на длительности не менее 5 сек)

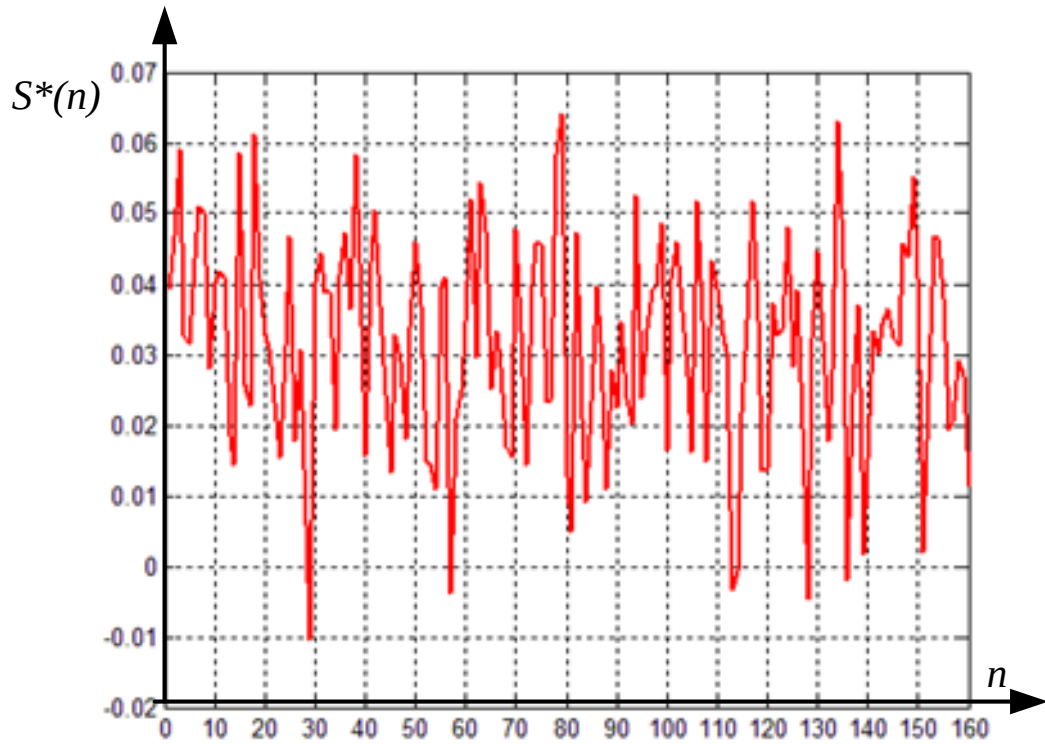


Рисунок 5.11 – Очищенный сегмент РС

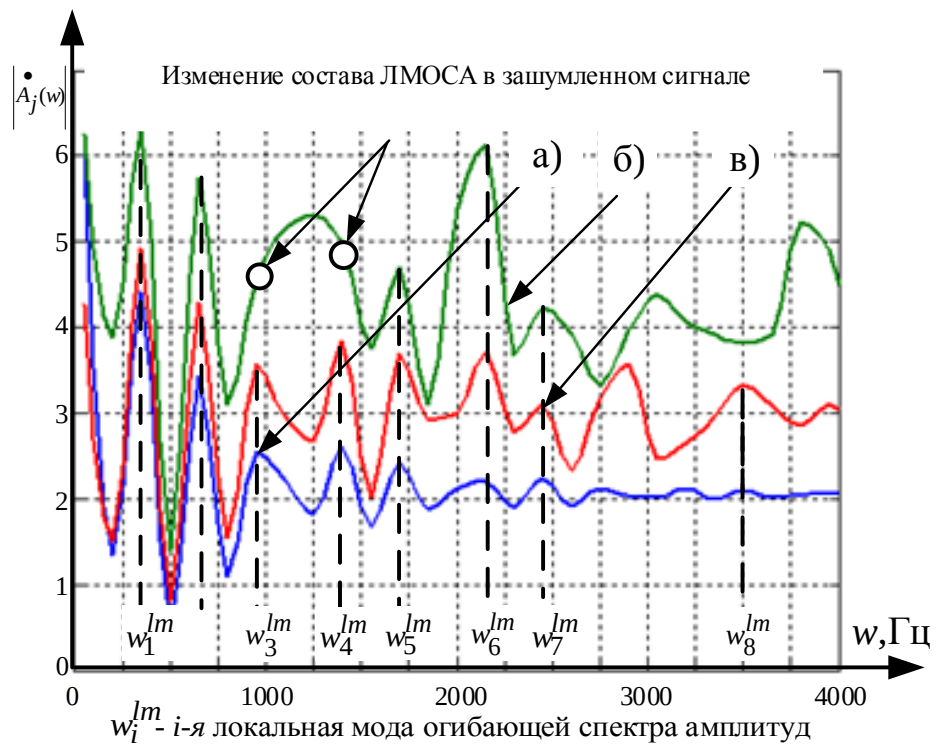


Рисунок 5.12 – Восстановление исходного состава ЛМОСА в очищенном РС:

- а) огибающая спектра амплитуд чистого РС;
- б) огибающая спектра амплитуд зашумленного РС;
- в) огибающая спектра амплитуд очищенного РС;

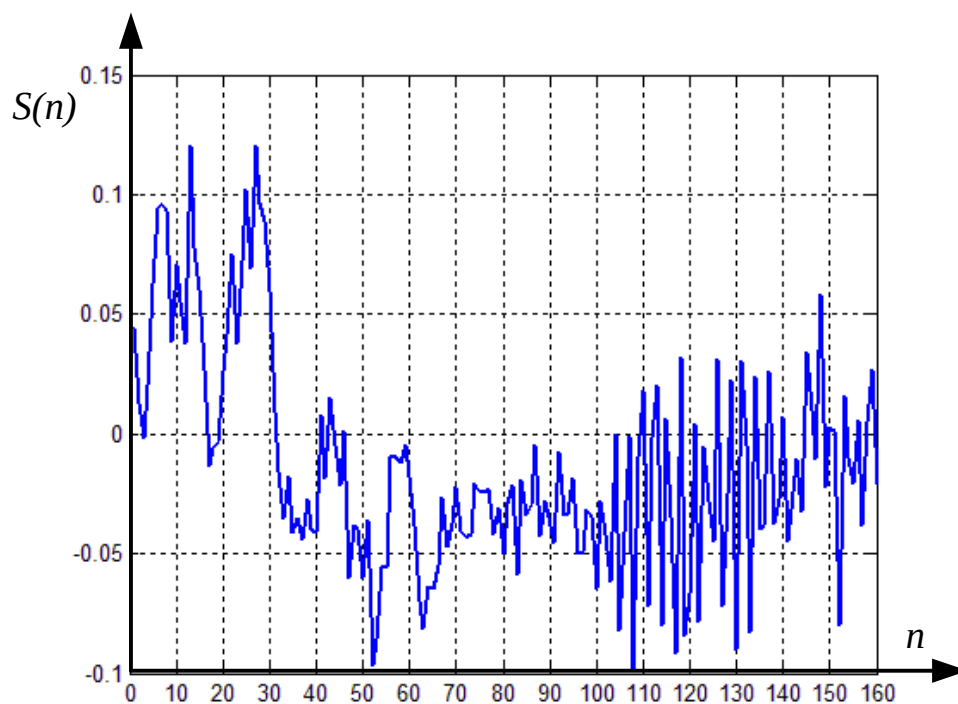


Рисунок 5.13 – Чистый сегмент РС, описывающий вокализованный элемент речи с относительно высокой энергетикой

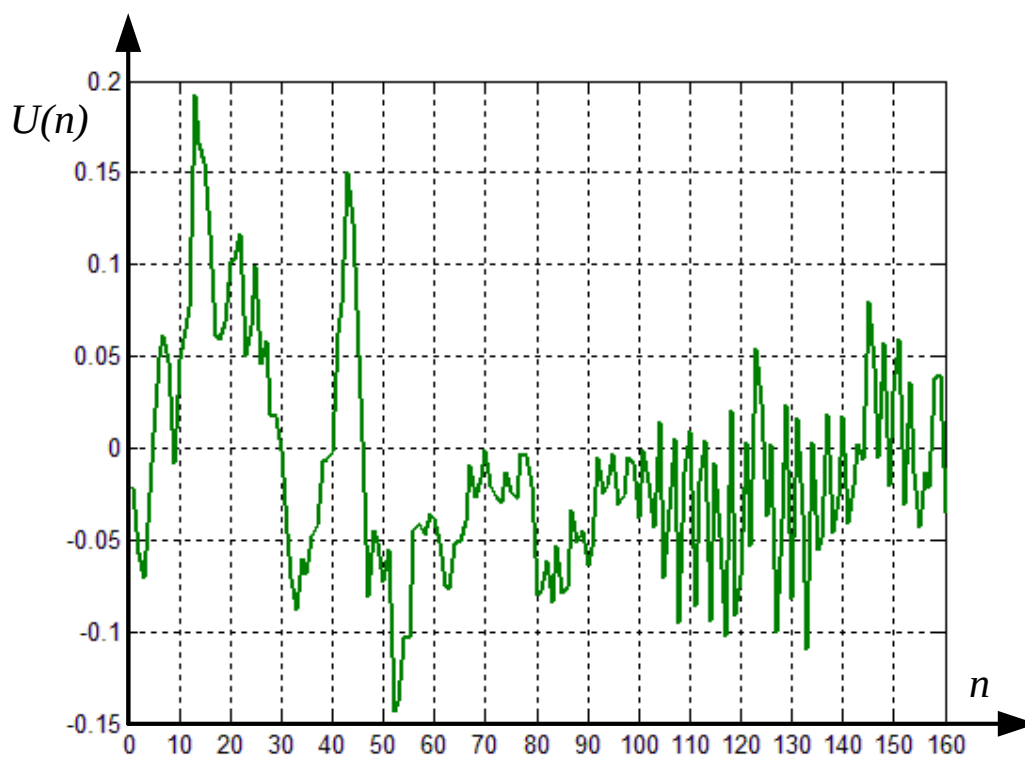


Рисунок 5.14 – Сегмент РС зашумленный шумом (шум двигателя) сегмент РС ОСШ = 0 дБ (измерение ОСШ осуществляется на длительности не менее 5 сек)

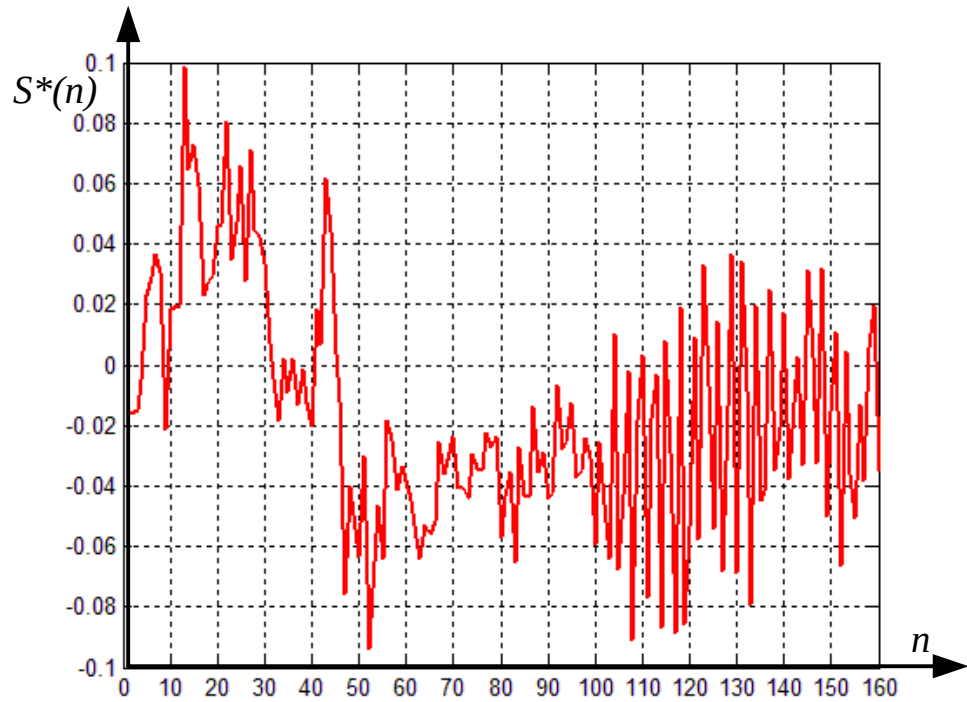


Рисунок 5.15 – Очищенный сегмент РС

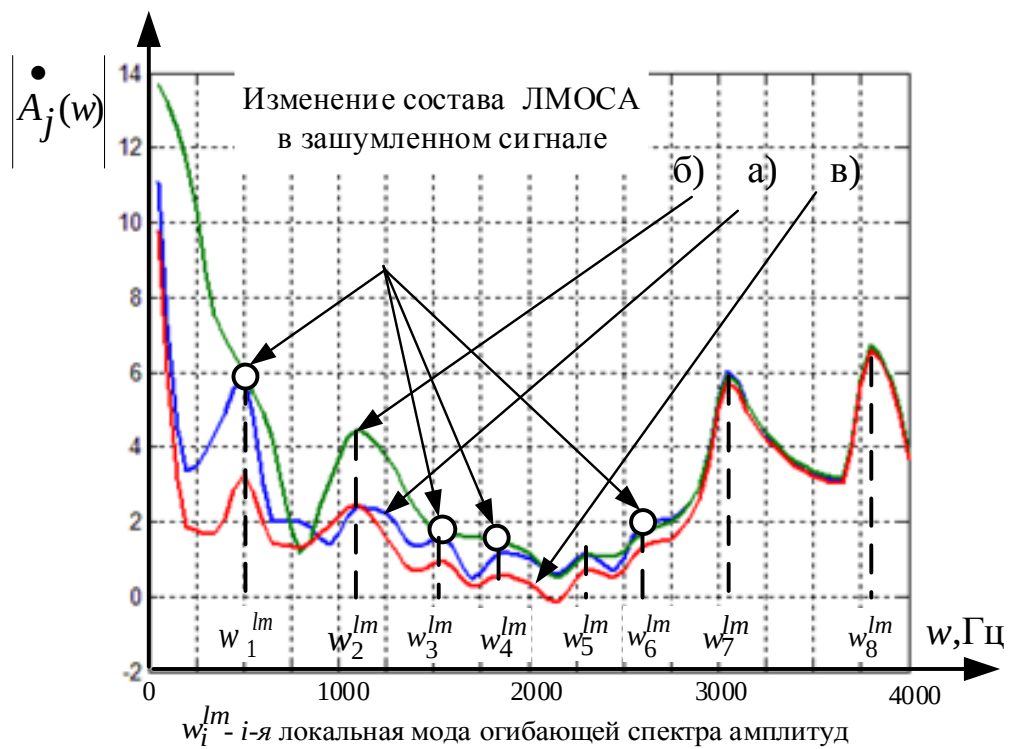


Рисунок 5.16 – Восстановление исходного состава ЛМОСА в очищенном РС:

- а) огибающая спектра амплитуд чистого РС;
- б) огибающая спектра амплитуд зашумленного РС;
- в) огибающая спектра амплитуд очищенного РС.

5.5 Выводы по пятому разделу

1. На основе представленной модели системы фильтрации речевого сигнала с применением полиспектрального анализа разработаны алгоритмы обработки РС с фиксированным и переменным сегментами анализа в условиях интенсивных акустических воздействий.

2. Разработанные алгоритмы фильтрации РС в УИАВ не требуют априорных сведений о характере и уровне акустического воздействия, не зависят от используемой частоты дискретизации и позволяют преодолеть частичную априорную неопределенность об обрабатываемом сигнале путем оценивания результатов анализа биспектров.

3. Сравнительная характеристика объективных и субъективных оценок на входе и выходе системы фильтрации позволяет сделать вывод о том, что разработанные алгоритмы позволяют значительно улучшить качество обрабатываемого речевого сигнала.

6 РАЗРАБОТКА АЛГОРИТМОВ ОБРАБОТКИ РЕЧЕВОГО СИГНАЛА В СИСТЕМАХ СВЯЗИ

6.1 Вводные замечания

В данном разделе рассмотрены подходы к определению характеристик речевого сигнала на сегменте анализа (подраздел 6.2) и способы классификации сегментов (подраздел 6.3), представлены технические решения по выбору длительности однородного сегмента анализа при его обработке на основе линейного предсказания (подраздел 6.4) и нахождения зависимостей параметров декомпозиции речевого сигнала при линейном предсказании (подраздел 6.5), выявлены особенности реализации и применения систем обработки речевого сигнала в средствах и комплексах инфокоммуникаций (подраздел 6.6).

6.2 Определение характеристик сегмента анализа речевого сигнала

Для реализации представленных в разделах 2 и 3 подходов к созданию систем обработки необходимо рассмотреть методы оценки характеристик РС. Для классификации сегментов РС на конечное число классов используются статистические и параметрические характеристики РС [44].

К статистическим характеристикам, определяемым на сегменте анализа и используемым в разрабатываемой системе обработки РС, относятся:

- математическое ожидание M ;
- средняя мощность (дисперсия) сигнала D ;
- функция автокорреляции B РС;
- функция плотности вероятности (ФПВ) $w(S)$ РС.

Математическое ожидание M и дисперсия D для речевого сегмента определяются формулами (6.1) и (6.2) соответственно [75]:

$$M = \sum_{i=1}^{N_{\text{сег}}} S(i)p_i, \quad (6.1)$$

$$D = \sum_{i=1}^{N_{\text{сег}}} (S(i) - M)^2 p_i, \quad (6.2)$$

где $N_{\text{сег}}$ – размер сегмента (количество отсчетов); p_i – вероятность появления отсчета $S(i)$ в сегменте.

При совместном изучении центрированных случайных величин вводится автокорреляционная функция (АКФ):

$$B(j) = \sum_{i=0}^{N_{\text{сег}}} S(i+1)S(i+j+1). \quad (6.3)$$

Выражение (6.3) справедливо при вычислении функции автокорреляции РС в сегменте для случая равновероятности появления отсчетов РС.

Вычисление функции автокорреляции является частью процедуры нахождения коэффициентов ЛП [116].

В [114] показано, что статистические свойства РС хорошо описываются четырьмя типами распределений со следующими функциями плотности вероятности: гауссовой (6.4), равномерной (6.5), лапласовой (6.6), гамма (6.7):

$$w(S) = \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{(S-M)^2}{2\sigma^2}\right)}; \quad (6.4)$$

$$w(S) = \begin{cases} \frac{1}{2\sqrt{3}\sigma}, & |S| \leq \sqrt{3}\sigma; \\ 0, & S > \sqrt{3}\sigma \end{cases}; \quad (6.5)$$

$$w(S) = \frac{1}{\sqrt{2}\sigma} e^{\left(-\frac{\sqrt{2}|S|}{\sigma}\right)}; \quad (6.6)$$

$$w(S) = \frac{\sqrt[4]{3}}{\sqrt{8|S|\pi\sigma}} e^{\left(-\frac{\sqrt{3}|S|}{2\sigma}\right)}. \quad (6.7)$$

где S – соответствующее значение отсчета РС; σ – среднеквадратическое отклонение сигнала (СКО), $\sigma = \sqrt{D}$.

К параметрическим характеристикам, которые могут быть использованы при разработке алгоритмов функционирования системы обработки РС, относятся:

- число переходов через ноль Z_n ;
- параметр сигнал тон/шум I (СТШ), вычисляемый на сегменте анализа, позволяющий разделить РС на 2 класса сегментов: шумовые и тоновые;
- частота ОТ $F_{от}$;
- число формантных областей ΔF_i на сегменте анализа.

Число переходов через ноль Z_n определяет шумовую составляющую РС. Это частота смены знака текущих отсчетов на сегменте анализа.

Для выделения параметра СТШ на сегменте анализа предполагается использовать метод, основанный на анализе АКФ [129], а также модифицированный метод Итакуры-Саито. Процесс нахождения I реализуется путем анализа АКФ сегмента сигнала длительностью 80 отсчетов и нахождения ее экстремума в заданной области определения [77, 95], при этом вводится следующее правило принятия решения:

$$\begin{aligned} B(\tau_i) \geq kB(0) &\rightarrow I - \text{тон}, \\ B(\tau_i) < kB(0) &\rightarrow I - \text{шум}. \end{aligned} \quad (6.8)$$

В работе используется значение $k = 0,5$, определяемое экспериментально.

Нахождение частоты ОТ производится для тоновых сегментов

$$F_{от} = \frac{1}{\tau_i}, \quad (6.9)$$

где τ_i – значение временного сдвига АКФ, при котором найден ее экстремум, удовлетворяющий (6.8).

Оценка параметра СТШ и частоты ОТ методом Итакуры-Саито заключается в фильтрации остатка линейного предсказания рекурсивным НЧ-фильтром 6-го порядка (прототип фильтра Баттерворта с частотой среза от 300 до 600 Гц) для уменьшения влияния второй формантной области на результаты вычислений [8, 134]. Данный фильтр выбран из соображений равномерности амплитудно-частотной и относительной линейности в рабочей полосе фазочастотной характеристик в полосе пропускания, а также приемлемой крутизны АЧХ в полосе расфильтровки.

Для оценки частоты ОТ используются одновременно два метода: метод анализа АКФ и модифицированный метод Итакуры-Саито. За частоту ОТ принимается среднее арифметическое значение, вычисленное с помощью этих двух методов:

$$F_{от} = 0,5(F_{от1} + F_{от2}).$$

Частота среза фильтра в методе Итакуры-Саито выбирается в зависимости от предварительно полученных данных о частоте ОТ на основе анализа АКФ речевого сегмента. Правила адаптивного изменения граничной частоты F_r пропускания фильтра в зависимости от частоты ОТ представлены в (таблице 6.1).

Значения граничных частот F_r , равные 300, 400, 600 Гц, выбраны путем моделирования на основе исследования распределения частоты ОТ для различных дикторов [16, 43] и влияния второй формантной области на результаты вычисления частоты ОТ.

Таблица 6.1 – Зависимость граничной частоты среза ФНЧ от частоты основного тона

$F_{от}, \text{ Гц}$	$F_r, \text{ Гц}$
< 130	300
130 – 200	400
> 200	600

Число формантных областей ΔF_i на сегменте анализа определяется посредством исследования статистически усредненной оценки СПМ РС, исследование формантной структуры речи представлено в [159, 177].

Оценка и использование данных параметров РС позволяет далее применять предлагаемые методы на практике.

6.3 Классификация сегментов речевого сигнала

Качественные характеристики разрабатываемой системы обработки с учетом зависимостей элементов декомпозиции РС зависят от выбранного алгоритма классификации речевых сегментов и алгоритмов кодирования, применяемых к каждому из сегментов в отдельности [28, 45].

Исследовались 3 альтернативных алгоритма классификации сегментов:

1. Классификатор, основанный на анализе статистических и параметрических характеристик речи [69, 188].

2. Классификатор, основанный на анализе функции распределения исследуемого сигнала [76].

3. Классификатор, основанный на цепях Маркова [76].

Классификатор первого типа основан на многоэтапном анализе сегмента. На первом этапе классификации сигнал делится на активные участки и паузы, критерием принятия решения служит соотношение (6.10).

$$P \leq P_0 \begin{cases} \text{да, } S - \text{пауза,} \\ \text{нет, } S - \text{речь,} \end{cases} \quad (6.10)$$

где $P = \frac{1}{N_{\text{сег}}} \sum_{i=1}^{N_{\text{сег}}} S_i^2$, P_0 – выбирается экспериментальным путем [159], $N_{\text{сег}}$ –

число отсчетов в анализируемом сегменте.

Определение P_0 заключается в разделении сегмента на паузы и активные участки с выбором начального уровня P_0 . Сформированный экспертом сигнал пауз подвергается прослушиванию и в нем определяется средняя мощность этих фрагментов для коррекции значения P_0 с учетом рекомендаций высказанных экспертом. Определение дисперсии сигнала паузы производится для большого числа дикторов с различными свойствами речи. Таким образом, за несколько итераций находится граничное значение дисперсии, которое в дальнейшем и принимается за величину P_0 .

Рассчитанное значение P_0 позволяет делить РС на 2 класса (пауза/речь) таким образом, что доля времени пауз в речи составляет около 50 %, что соответствует результатам исследований речи другими авторами [186].

На втором этапе классификации сегменты активной речи делятся на 4 типа: тоновые, шумовые, переходные 1 типа и переходные 2 типа. Для такого деления вычисляется параметр СТШ I_i и частота ОТ $F_{от}$ на сегменте анализа.

Вычисления параметра СТШ и частоты ОТ производятся совместно на основе анализа АКФ РС и метода Итакуры-Саито. Сочетание двух методов в совокупности уменьшает вероятность ошибки в классификации сегментов речи. Целесообразность выбора данных методов для выделения параметра СТШ и частоты ОТ объясняется уменьшением общей вычислительной сложности алгоритма кодирования за счет повторного использования результатов, полученных в ходе вычисления данных параметров, на последующих этапах обработки РС.

Правила принятия решения о типе сегмента речи сформулированы следующим образом:

- тоновые сегменты: метод анализа АКФ определяет сегмент как тоновой, метод Итакуры-Саито определяет сегмент как тоновой;
- шумовые сегменты: метод анализа АКФ определяет сегмент как шумовой, метод Итакуры-Саито определяет сегмент как шумовой;
- переходные сегменты первого типа: метод анализа АКФ определяет сегмент как шумовой, метод Итакуры-Саито определяет сегмент как тоновой;
- переходные сегменты второго типа: метод анализа АКФ определяет сегмент как тоновой, метод Итакуры-Саито определяет сегмент как шумовой.

На третьем этапе все тоновые и переходные сегменты РС делятся на мужские и женские по правилу: если $F_{от} > F_{гр}$, то сегмент «женский», иначе $F_{от} \leq F_{гр}$, то сегмент «мужской». Граничная частота ОТ, разделяющая мужские и женские голоса, определена экспериментально $F_{гр} = 180$ Гц и соответствует результатам исследований других авторов [186]. Таким образом, с помощью трехэтапного классификатора активные сегменты РС будут делиться на 7 типов, то есть в цифровом потоке будет задействовано 3 бита для кодирования структуры кодека.

Алгоритм классификации речевых сегментов на конечное число классов с использованием характеристик РС представлен на (рисунке 6.1), а варианты классифицированных сегментов показаны на (рисунке 6.2).

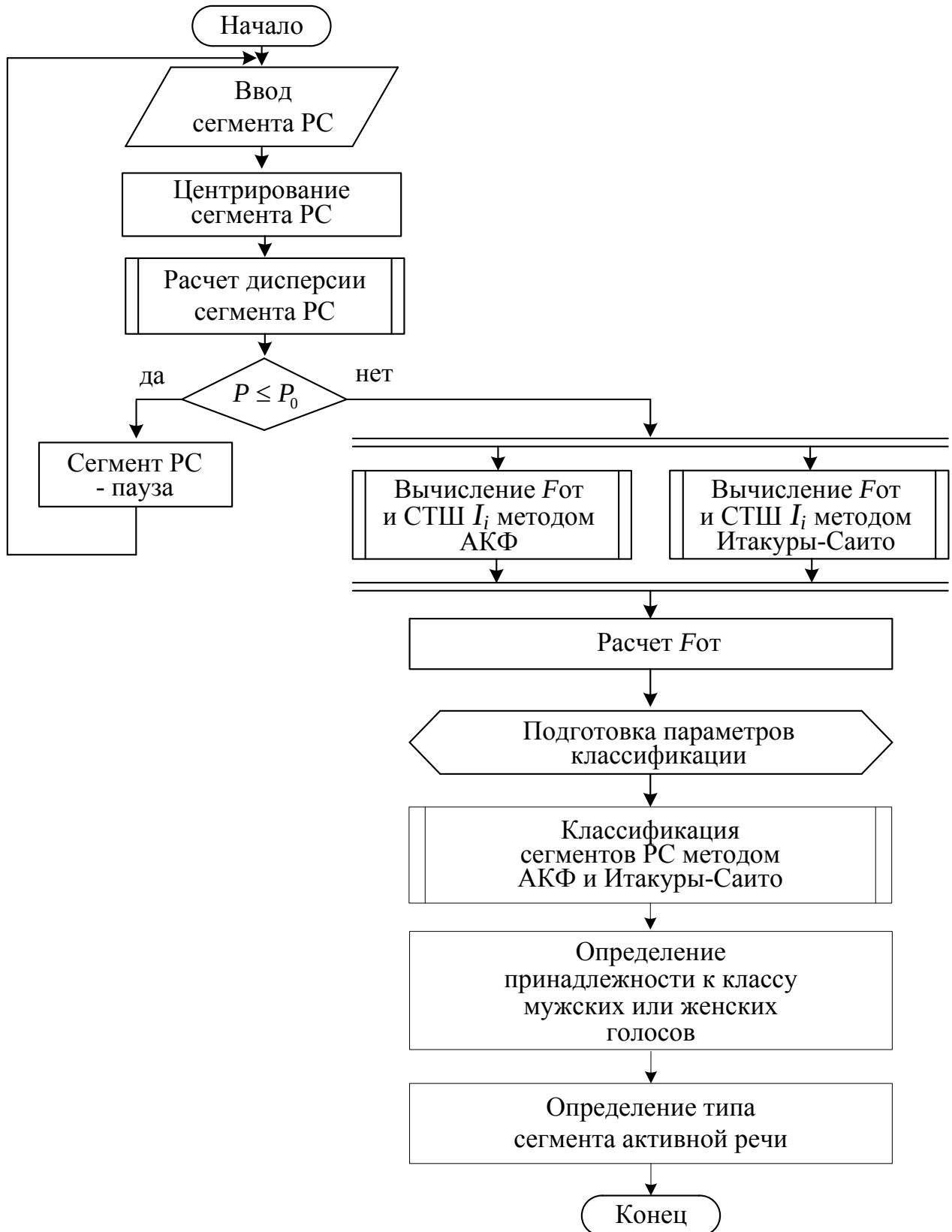


Рисунок 6.1 – Блок-схема алгоритма классификации сегментов РС

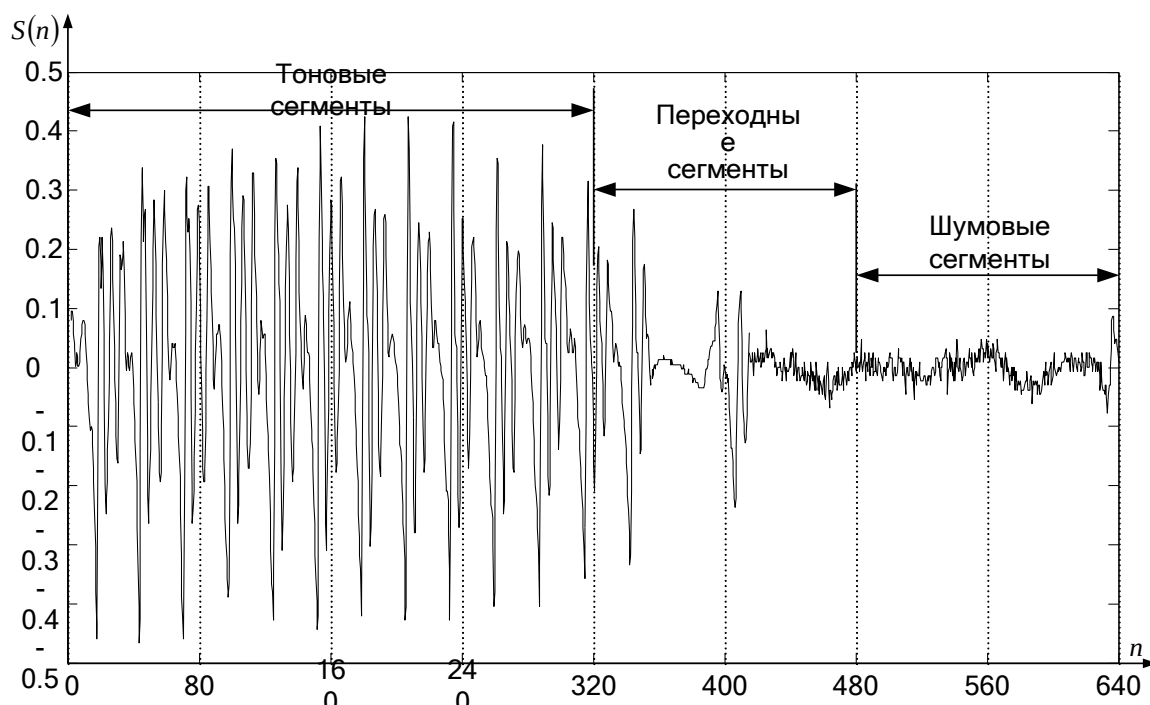


Рисунок 6.2 – Виды речевых сегментов во временной области

Принцип построения классификатора второго типа состоит в использовании четырех основных функций плотности вероятности, представленных в п. 4.4, законы распределения которых описывают статистические свойства РС. Такой классификатор строится на основе критерия согласия с четырьмя возможными аналитическими моделями функций плотности вероятности (6.4)–(6.7). Данный классификатор дополнительно может использовать и такие характеристики, как мощность сигнала на участке анализа, параметр тон/шум и частоту ОТ, функцию корреляции сигнала на сегменте анализа (6.3). Использование данного подхода не всегда приемлемо, так как чаще для дальнейшей обработки РС используются параметры, вычисленные при реализации классификатора первого типа.

Классификатор третьего типа основан на использовании цепей Маркова для разделения сегментов речи. Одной из самых распространенных и простых моделей на основе марковских цепей является модель перехода с тремя состояниями: 1 – пауза, 2 – вокализованный сегмент, 3 – шумовой сегмент. Данная модель описана в [188].

Матрица переходных вероятностей P в данном случае будет иметь вид (6.11),

где индексы p,v,n – соответствуют паузе, тоновому и шумовому сегменту соответственно.

$$R = \begin{vmatrix} p_{00} & p_{01} & p_{02} \\ p_{10} & p_{11} & p_{12} \\ p_{20} & p_{21} & p_{22} \end{vmatrix} = \begin{vmatrix} p_{pp} & p_{pv} & p_{pn} \\ p_{vp} & p_{vv} & p_{vn} \\ p_{np} & p_{nv} & p_{nn} \end{vmatrix}. \quad (6.11)$$

Матрица финальных вероятностей в данном случае имеет вид (6.12):

$$R = \begin{vmatrix} P_0 \\ P_1 \\ P_2 \end{vmatrix} = \begin{vmatrix} P_p \\ P_v \\ P_n \end{vmatrix}. \quad (6.12)$$

Финальные вероятности состояний находятся из системы уравнений (6.12).

$$\sum_{i=0}^{K-1} P_i = 1; \quad \sum_{j=1}^{K-1} P_j p_{jk} = P_k. \quad (6.13)$$

Учитывая порядок уравнения, равный трем, найдем финальные вероятности состояний.

$$P_0 = \frac{A}{A+B+C}; \quad P_1 = \frac{B}{A+B+C}; \quad P_2 = \frac{C}{A+B+C}; \quad (6.14)$$

где $A = p_{10} p_{20} + p_{12} p_{20} + p_{21} p_{10}$; $B = p_{01} p_{21} + p_{02} p_{21} + p_{20} p_{01}$;

$C = p_{02} p_{12} + p_{01} p_{12} + p_{10} p_{02}$.

Данные формулы описывают модель переходного процесса с тремя состояниями: пауза, тоновой сегмент, шумовой сегмент. На основе такой модели можно с высокой долей вероятности классифицировать приходящие речевые сегменты.

Однако данный подход не отражает всех характеристик речевого сегмента и целесообразнее в данном случае ввести шестипараметрическую модель РС. Шестью состояниями цепи в данном случае будут являться следующие события: пауза,

вокализованный сегмент, шумовой сегмент с малой энергией, шумовой сегмент с высокой энергией, вокализованный сегмент с малым коэффициентом сложности, вокализованный сегмент с высоким коэффициентом сложности[6, 28]. Матрица переходных вероятностей в данном случае усложняется, появляется возможность более точного учета свойств речевого сегмента. Однако на практике получить точное статистическое описание и достаточно достоверную матрицу переходных вероятностей сложно. С точки зрения простоты реализации и качественных показателей функционирования в лучшую сторону выделяется классификатор первого типа, поэтому он выбран в качестве классификатора сегментов РС для дальнейшего использования при обработке РС.

6.4 Определение длительности сегмента анализа речевого сигнала

Процедура выделения однородных сегментов при анализе речи обеспечивает выделение сегментов кратных периоду OT , что приводит к увеличению точности параметризации за счет уменьшения влияния явления растекания спектра [4, 37, 44, 55]. На рисунке 2.5 была приведена иллюстрация метода динамического выделения сегментов анализа РС.

Выбор длины данного сегмента является весьма важной задачей. Его увеличение приводит к уменьшению скорости передачи в КС, а сокращение – к повышению качественных характеристик синтезируемого сигнала, так как сигнал становится близким к стационарному, и уменьшению времени задержки сигнала на обработку. Анализ речевого сигнала на фиксировано выбранном однородном квазистационарном сегменте является достаточно грубым допущением, так как за пределами сегмента сигнал представляется равным нулю, что не соответствует действительности и приводит к появлению искажений на стыках сегментов при их анализе и кодировании, а также искажений в восприятии синтезированного РС на приеме. При этом на вокализованных участках речи длина сегмента стационарности может быть увеличена благодаря линейному характеру образования РС на этом участке, а на шумоподобных участках желательно ее уменьшать, так как речевой

сигнал в данном случае имеет нестационарные свойства [161, 186]. Деление РС на вокализованные и шумоподобные (невокализованные) сегменты широко используется в технике эффективного речевого кодирования [188]. В результате применения предлагаемого способа налицо снижение требуемой скорости передачи при сохранении качественных показателей синтезированного РС.

Рассмотрим сущность предлагаемого метода. В качестве входного сигнала используется оцифрованный РС, полученный на выходе линейной ИКМ с параметрами определяемыми стандартом G.711 [167].

Данный РС разделяется на сегменты длительностью по 20 мс. Если принимается решение о том, что сигнал является активной речью, то выделяется переход сигнала через нулевое значение, и от положения отсчета со значением, наиболее близким к нулю, выбирается длина сегмента, соответствующая 20 мс, и рассчитывается значение частоты ОТ и сигнала тон-шум. Далее, если принимается решение о вокализованности анализируемого сигнала, то увеличивается длительность сегмента квазистационарности на количество отсчетов, кратное периоду ОТ, но не более чем на 60 мс, с обязательной проверкой на вокализованность следующих сегментов по 20 мс. Если принимается решение о шумоподобности следующего сегмента, то граница сегмента анализа выбирается кратной количеству отсчетов на периоде ОТ, но не более половины следующего сегмента длительностью 20 мс. Но если принимается решение о шумоподобности анализируемого сегмента, то длина сегмента анализа уменьшается, при этом граница сегмента формируется на значении близком нулю и кратном вычисленному периоду ОТ на предыдущем сегменте.

При таком подходе начальный и конечный отсчеты во вновь сформированном сегменте будут иметь значения близкие нулю, что значительно уменьшит возможные искажения на стыках сегментов [130, 133].

В большинстве вокодеров с линейным предсказанием одним из параметров, входящих в кадр передачи, является частота ОТ говорящего.

Таким образом, использование его в качестве одного из критериев формирования границ сегмента квазистационарности РС не приводит к повышению скорости передачи в таких вокодерах, а существенно ее сокращает. При этом повышается

качество синтезированной речи на выходе вокодера с ЛП, так как отсчеты на стыках сегментов квазистационарности имеют практически нулевую энергию. В некоторых методах ЛП [188] частота ОТ говорящего не передается в кадре передачи. Для таких методов данный параметр будет дополнительным, что снижает эффект от внедрения предлагаемого метода, однако общий выигрыш от его использования будет достаточно велик. Использование предлагаемого способа для выделения сегментов квазистационарности при анализе РС в вокодерах с ЛП рационально для класса систем кодирования РС с переменной скоростью передачи данных по КС. Такой подход позволяет снизить требуемую пропускную способность КС для функционирования систем с ЛП, а качественные показатели синтезированного сигнала значительно повышаются. Алгоритм выделения сегментов при анализе РС в вокодерах с ЛП представлен на (рисунке 6.3). На (рисунке 6.4) показана функциональная схема устройства, реализующего такой алгоритм. В функциональной схеме устройства (рисунок 6.4) применены: блок 1 приема сегмента цифрового РС, блок 2 анализа текущего сегмента на наличие активной речи, блок 3 выделения перехода огибающей РС через нулевое значение, блок 4 поиска отсчета со значением наиболее близким к нулевому, блок 5 выделения сигнала тон-шум и вычисления значения частоты ОТ, блок 6 увеличения длительности сегмента, блок 7 уменьшения длительности сегмента, получатель 8 сегмента цифрового РС.

К достоинствам предлагаемого технического решения относится тот факт, что изменение длительности сегментов квазистационарности при анализе РС в вокодерах с ЛП дает возможность уменьшить среднюю скорость передачи в липредерах, функционирующих с переменной скоростью передачи данных. Применение такого способа позволит понизить среднюю скорость передачи данных в КС по сравнению с известными решениями, в которых используется фиксированный сегмент анализа РС.

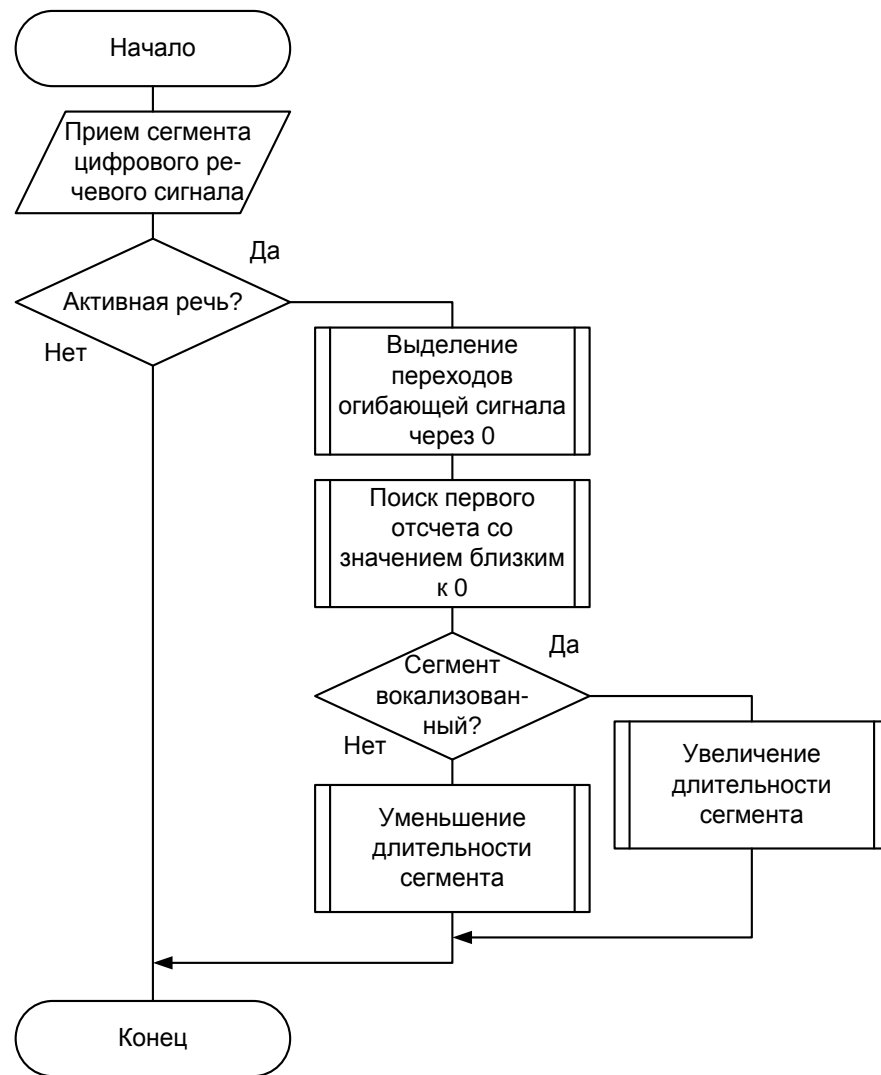


Рисунок 6.3 – Алгоритм выделения однородных сегментов анализа речевого сигнала в вокодерах с линейным предсказанием

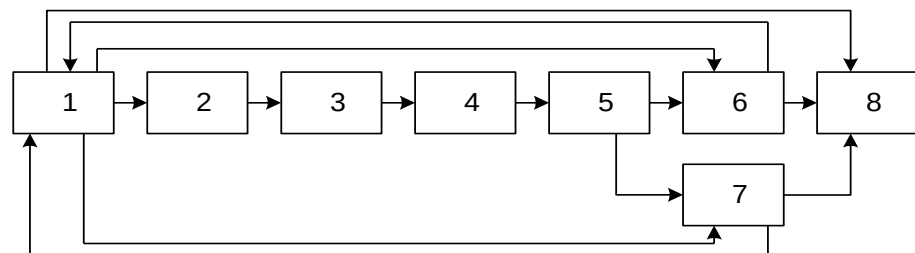


Рисунок 6.4 – Функциональная схема устройства, выделяющего однородные сегменты при анализе речевого сигнала в липредерах

Использование в математической модели ЛП фиксированных длин сегментов анализа приводит к возникновению скачков огибающей синтезированного РС на их стыках, что связано с дискретным изменением коэффициентов формирующей

модели ЛП. Следует также учитывать тот факт, что длительность сегментов, используемая при ЛП, гораздо меньше длины участка установившегося звука (фонемы) речи, что подтверждается данными, представленными в [69, 186].

Исследование природы формирования акустического сигнала в речевом аппарате человека показало наличие подавляющего числа вокализованных участков при его анализе. Вокализованный РС характеризуется формантной структурой, при этом основная энергия в нем сосредоточена в области первой-второй форманты, что определяет использование соответствующего частотного диапазона (таблица 6.2). Таким образом, в основе метода выделения переменных длин однородных сегментов анализа речи в вокодерах с линейным предсказанием лежит отслеживание факта перехода огибающей РС во временной области через ноль при сохранении знака конечной разности первого порядка в начале и конце анализируемого сегмента.

Таблица 6.2 – Распределение формант по диапазону частот

Распределение формант по частоте и амплитуде	Диапазон частот для формант вокализованных звуков:	150 – 900 Гц 550 – 2800 Гц 1500 – 3400 Гц	
	- первая форманта - вторая форманта - третья форманта		
	Диапазон частот для формант невокализованных звуков, Гц:	Для естественной речи	Для телефонного канала
	- первая форманта - вторая форманта - антиформанта	1000 – 3500 2500 – 6000 1500 - 4000	1000 – 3400 2500 – 3400 1500 - 3400

Такой подход позволяет уменьшить влияние граничных эффектов, связанных с переходными процессами в синтезирующем фильтре, а также эффект растекания спектра. Нахождение областей локализации перехода связано с оценкой параметра частоты ОТ говорящего при учете корреляционных зависимостей в речи. Для их учета в исследуемой модели используются критерии однородности выборок, определяющие ее принадлежность к генеральной совокупности [87, 95]. Для различных элементов формируемой речи длительности устоявшегося процесса значительно

превышают 10...30 мс, при этом длительность отдельных звуков речи (фонем) составляет от 20 до 350 мс [69]. Гласные имеют большую длительность, чем согласные. Темп речи может изменяться в широких пределах, причем длительность гласных звуков изменяется в большей степени. Максимально возможный интервал одновременно анализируемых данных определяется рекомендацией G.114 Международного союза электросвязи и составляет 60-80 мс, что связано с требованиями по задержке РС при его передаче. Таким образом, если на протяжении данного сегмента не происходит смены однородности при формировании РС, то новый анализ начинается при использовании данных о выделенных параметрах предыдущего сегмента. Использование данного метода обработки речевых данных позволяет выделять однородные сегменты, имеющие одинаковую природу формирования звука в речевом аппарате человека. Увеличение длительности кратковременного анализа приведет к тому, что параметры формирующей (передаточной) функции системы обработки и сигнала возбуждения будут сохраняться на всем протяжении анализа, что эквивалентно сокращению средней скорости передачи речевого сигнала. Так как РС является случайным нестационарным процессом, для которого характерны изменения параметров ОТ, значений коэффициентов, характеризующих передаточную функцию голосового тракта и вида сигнала возбуждения, то повышение пропускной способности сетей связи с коммутацией пакетов при их совместном использовании с приложениями IP – телефонии и передачи данных возможно за счет реализации алгоритмов речевого кодирования с переменной скоростью, такой подход представлен в [138].

В качестве примера рассмотрим вокализованный сегмент данных на интервале нескольких периодов ОТ (рис.2.5). Одним из принципиальных моментов при формировании конечной границы интервала анализа речевого сегмента является вычисление периода ОТ и изменение ряда значений интервалов корреляции. Его вычисление показано в [116, 129]. При этом границы сегмента формируются на основании выражений (2.32-2.34).

Использование выражений (2.32)–(2.34) позволяет получать сегменты анализа, начало и окончание которых имеют одинаковые знаки конечной разности

первого порядка. При этом с высокой вероятностью можно утверждать, что начальный и конечный отсчеты во вновь сформированном сегменте будут иметь значения близкие к переходу через ноль, что значительно уменьшит возможные искажения на стыках сегментов. Предлагаемое выделение параметров модели РС на однородных сегментах достаточно хорошо соотносится с квазистационарным характером РС на временных интервалах, соответствующих режиму установившихся звуков. Развитие данного метода на основе использования последовательного статистического анализа, раскрытого в п.2.3, показано на (рисунке 6.5), где представлена блок-схема алгоритма его реализации.

В данном способе выделение сегментов обработки речи реализуют на основе последовательного статистического анализа. На выходе аналого-цифрового преобразователя с частотой дискретизации 8 кГц и 256 уровнями квантования в РС выделяют участок анализа длительностью 20 мс. После этого оценивают кривую плотности вероятности методом ядерного сглаживания Парзена [71]. При этом количество окон сглаживания определяют с помощью информационного критерия Акаике [191]. Далее увеличивают сегмент анализа на 5мс, после чего рассчитывают значение логарифма правдоподобия для вновь полученной выборки по кривой плотности для начального сегмента 20мс. Также имеется возможность увеличения длительности сегмента анализа кратно значению вычисленного периода ОТ. На следующем шаге производят ядерную оценку плотности для вновь полученного набора отсчётов и по ней вычисляют логарифм правдоподобия. Используя критерий Вальда, определяют логарифм отношения правдоподобия, как разность уже вычисленных логарифмов. При значении логарифма отношения соответствующему «абсолютно эффективной» области $C_{эф}^{abc}$ увеличивают сегмент анализа ещё на 5 мс и вновь повторяют вышеописанные операции.

Увеличение продолжается до достижения размера сегмента анализа предельного значения в 80 мс, или до перехода значения логарифма отношения вероятностей в «неэффективную» область $C_{н.эф}$. При этом за конечную границу сегмента, принимают ту, при которой значение логарифма отношения принадлежало $C_{эф}^{abc}$, т.е. на предыдущем этапе наблюдения.

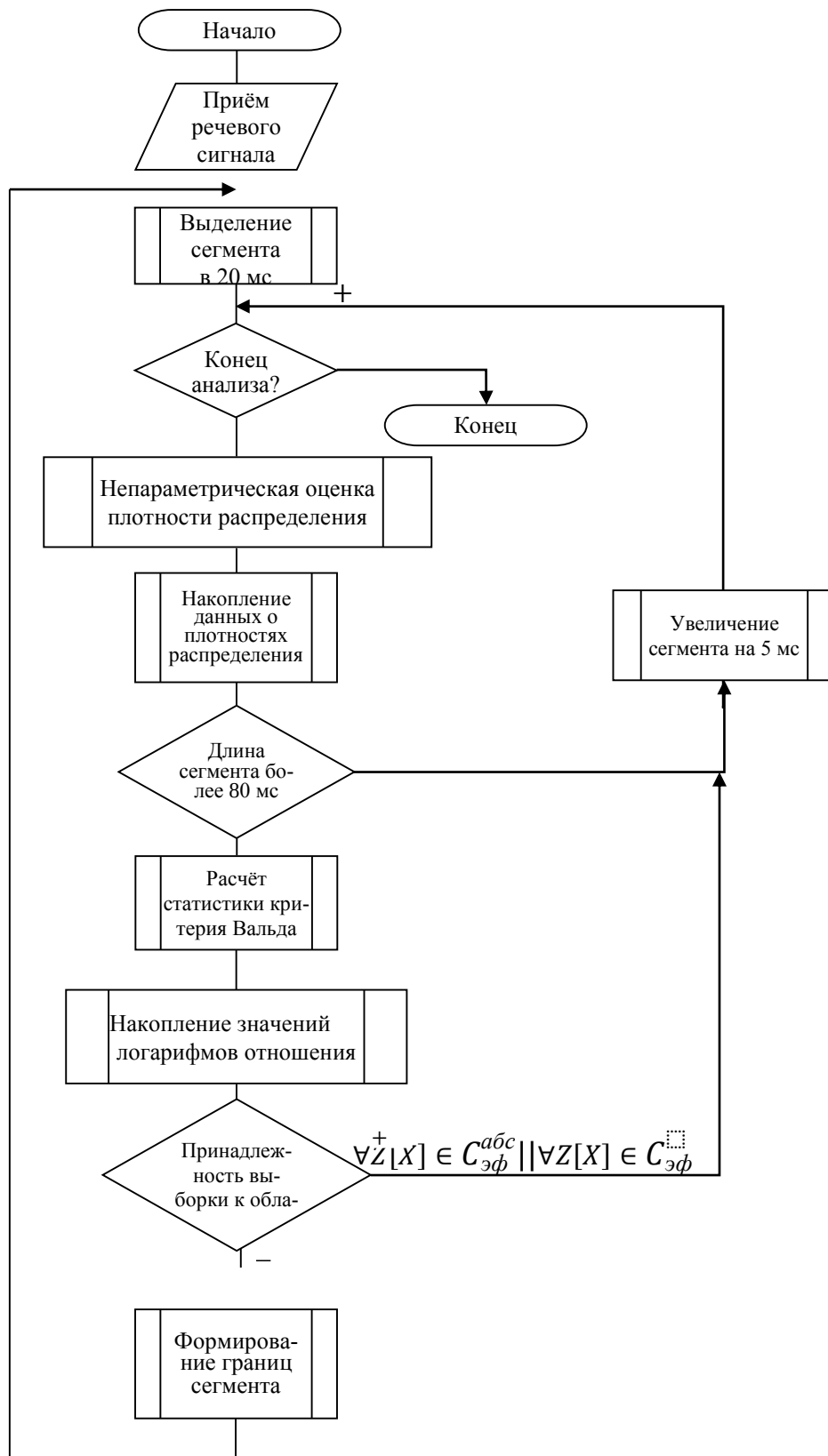


Рисунок 6.5 – Блок-схема алгоритма выделения сегментов обработки речи на основе последовательного статистического анализа

Если после первого увеличения начального сегмента анализа на 5 мс, значение логарифма отношения принадлежит «эффективной» области $C_{эф}$, то увеличение сегмента продолжается до перехода этого значения в абсолютно «эффективную» область $C_{эф}^{abs}$, либо до достижения размера сегмента анализа предельных значений в 80 мс, при этом за конечную границу принимается, та которая получена при этом переходе, т.е. на текущем, а не предыдущем этапе наблюдения. При значении логарифма, после первого увеличения сегмента, принадлежащем «неэффективной» $C_{н.эф}$ области, процесс увеличения прекращается и сегмент остаётся в своих первоначальных границах, определяемых каждый раз на основе текущего анализа.

Такой подход показывает пути уменьшения объема данных при кодировании РС, что приводит к получению необходимого и достаточного количества отсчетов для его качественного представления при обработке. К достоинствам следует отнести тот факт, что уменьшение объема данных при низкоскоростном кодировании и обработке речи приведет к снижению средней скорости передачи РС по каналам цифровой связи, а также уменьшит количество требуемых вычислений при реализации процедуры кодирования.

Были проведены экспериментальные исследования для выявления возможности применения предлагаемого способа, которые показали уменьшение объема анализируемых данных РС при низкоскоростном кодировании в среднем на 20% при описании формирующей (передаточной функции) голосового тракта на однородных сегментах при незначительном ухудшении качественных показателей синтезированного РС. Анализ (таблицы 3.1) дает основание утверждать, что в низкоскоростных методах кодирования РС около 25% информационного ресурса кадра передачи используется для передачи данных о параметрах формирующей функции, около 50% в РС занимают паузы, а 70% активной речи занимают вокализованные участки [69, 186]. Таким образом, выделение однородных участков РС на длительности нескольких классических сегментов анализа дает возможность до 20% уменьшать среднюю скорость передачи данных в КС, за счет сохранения информа-

ции о параметрах формирующей модели голосового тракта на них. Данные результаты являются следствием внедрения динамического изменения границ сегмента анализа и расчета значений параметров, описывающих передаточную функцию голосового тракта на этих сегментах.

Вычисляемые в ходе исследований значения ЛСЧ позволяют утверждать о высокой корреляции между параметрами, описывающими значения максимумов амплитудно-частотной характеристики речевого тракта на близлежащих однородных сегментах РС, что используется при его обработке в системах низкоскоростного кодирования.

Определение соотношений в выделяемом ресурсе кадра передачи между параметрами, описывающими передаточную функцию голосового тракта, и сигналом возбуждения является достаточно актуальной задачей, так как это позволит управлять скоростью передачи кодека речи, подстраивая его под требуемые качественные характеристики синтезируемого сигнала. Однако данная задача является весьма сложно формализуемой, так как нет возможности априорно указать вид и распределение речевых данных. Предположение же о нормальном законе распределения речевых данных часто встречающееся в научной литературе при математических выкладках является обоснованным допущением и приводит к некоторому расхождению аналитически и практически получаемых результатов при моделировании. Одним из возможных направлений решения данной задачи является выявление и нахождение зависимостей между параметрами передаточной функции синтезирующего фильтра и соответствующим им сигналом возбуждения. Существование данных зависимостей объясняется переходом от авторегрессионной модели параметрического ЦСА к модели ЛП. При этом, в отличие от классической постановки задачи ЦСА, используются фиксированные порядки формирующей модели липредера, что порождает наличие остаточной корреляции в сигнале ошибки. Использование данных зависимостей при учете переменной длины сегмента анализа дает возможность сократить среднюю скорость передачи данных в КС.

6.5 Определение взаимных связей элементов декомпозиции речевого сигнала

6.5.1 Процедура векторного квантования элементов декомпозиции речевого сигнала

При декомпозиции речи на основе ЛП формируются пространства векторов данных элементов фиксированной мощности. В основе построения данных пространств лежит процедура ВК. ВК является важной макропроцедурой подавляющего большинства современных алгоритмов кодирования РС [188].

Когда каждый отсчет РС квантуется отдельно, имеет место скалярное квантование. Когда совокупность отсчетов квантуется совместно как единый вектор, говорят о векторном (блочном) квантовании.

Процесс ВК заключается в следующем. Предполагается, что $\vec{S} = [s_1, s_2 \dots s_N]$ представляет собой N-мерный вектор параметров РС, компоненты которого $\{s_k, 1 \leq k \leq N\}$ - действительные случайные величины. При ВК входной вектор $\vec{S}$ отображается в N-мерный действительный вектор $\vec{S}'$ с определенными дискретными значениями своих координат. Этот процесс называется квантованием $\vec{S}$ в $\vec{S}'$, а $\vec{S}'$ представляет собой квантованное, значение $\vec{S}$. Процесс векторного квантования можно описать отображением:

$$\vec{S}' = q(\vec{S}), \quad (6.15)$$

где $q(*)$ — оператор квантования, $\vec{S}'$ называется также преобразованным вектором или выходным вектором соответствующим $\vec{S}$. Обычно $\vec{S}'$ принимает одно значение из ограниченного множества:

$$S' = \{ \vec{S}'_i, 1 \leq i \leq L \}, \text{ где } \vec{S}'_i = [s'_{i1}, s'_{i2} \dots s'_{iN}]^T. \quad (6.16)$$

Множество S' называется КК преобразования или просто КК. L — размер КК, а $\vec{S}'_i$ — множество кодовых векторов. Для построения такой КК N -мерное пространство случайного вектора $\vec{X}$ разделяется на L областей или ячеек $\{C_i, 1 \leq i \leq L\}$ и с каждой ячейкой C_i связывают вектор $\vec{S}'_i$. Тогда квантователь назначает кодовый вектор $\vec{S}'_i$, если $\vec{S}$ лежит в C_i :

$$q(\vec{S}) = \vec{S}'_i, \text{ если } \vec{S} \in C_i. \quad (6.17)$$

Блок-схема алгоритма работы ВК представлена на (рисунке 5.22). Вариант построения КК известен, как процесс обучения или заполнения КК [114, 188]. Данная процедура является неотъемлемой частью процедуры ВК, так как корректное формирование КК обуславливает устойчивую работу ВК при заданном минимуме искажений входного вектора. На (рисунке 6.6) приведен пример разделения двумерного пространства ($N=2$) при ВК. Область, окруженная линиями, представляет собой ячейку C_i , жирной линией на рисунке ограничена область изменения вектора $\vec{C}$. Любой входной вектор $\vec{S}$ лежащий в ячейке C_i , отображается в один из векторов $\vec{S}'_i$. Положения кодовых векторов других ячеек обозначены точками.

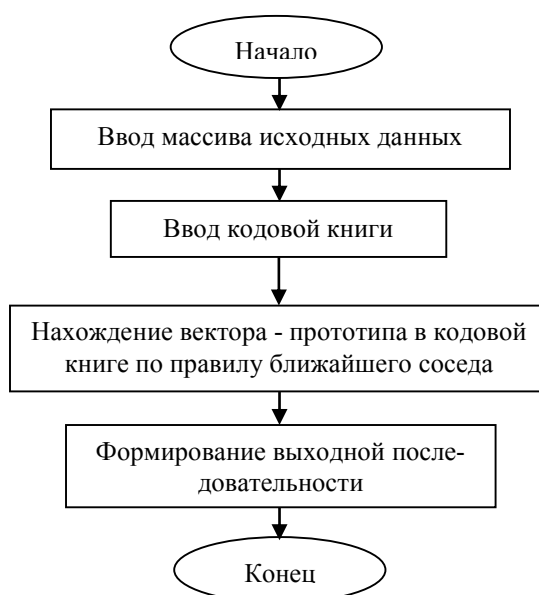


Рисунок 6.6 — Блок-схема алгоритма векторного квантования

Таким образом, вместо отдельного скалярного квантования элементов декомпозиции РС каждой координаты векторов конкретного параметра, например, $\vec{S} = (s_1, s_2, s_3, \dots, s_n, \dots, s_N)$ применяется совместное одновременное квантование всех координат $(s_1, s_2, s_3, \dots, s_n, \dots, s_N)$. При квантовании текущего значения случайного вектора $\vec{S}$ по КС в виде двоичной комбинации передается лишь номер i -го кластера, к которому принадлежит квантуемый вектор $\vec{S}$ (номер центроида, хранящий координаты $\vec{S}'$). На рисунке 6.7 представлена графическая иллюстрация ВК для двумерного случая ($Q=2$), жирной линией на рисунке ограничена область изменения вектора $\vec{S}$.

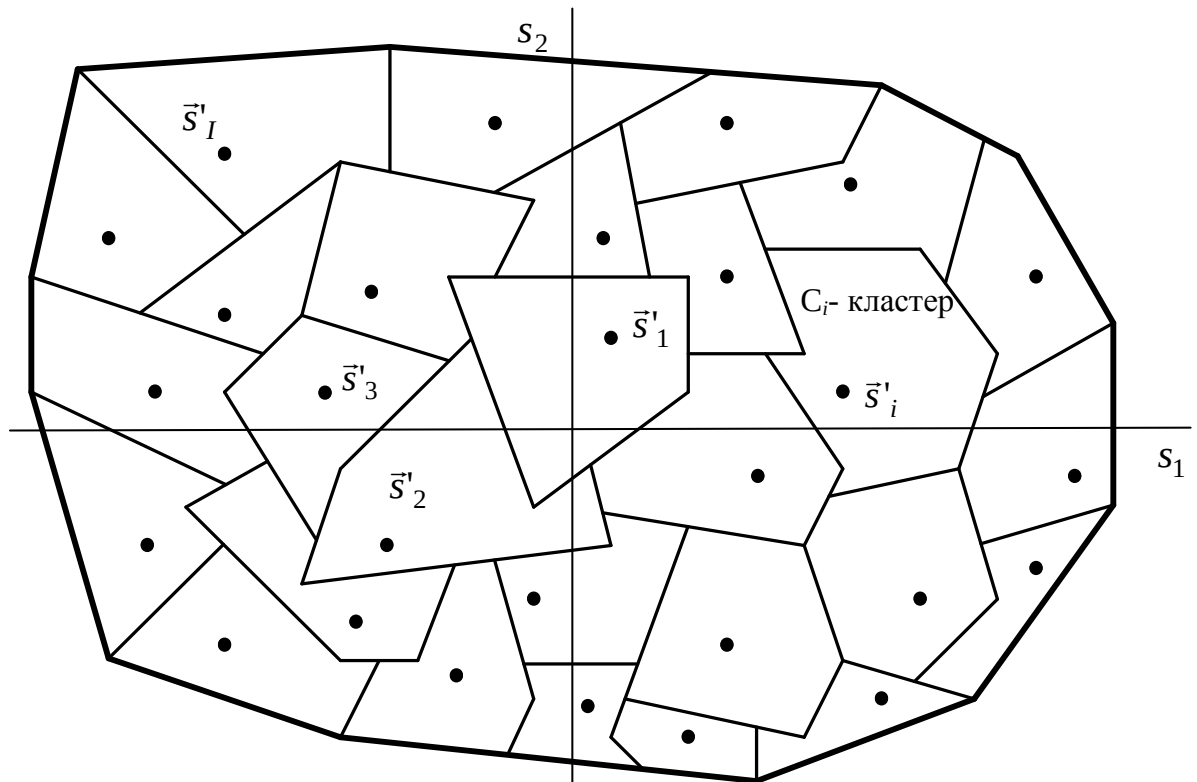


Рис. 6.7 – Графическое представление процедуры векторного (двумерного) квантования

При квантовании $\vec{S}$ в $\vec{S}'$ возникает ошибка квантования. Отклонение $\vec{S}$ от $\vec{S}'$ может быть определено мерой искажения $d(\vec{S}, \vec{S}')$, называемой также мерой расхождения [114]. В случае передачи векторов $\vec{S}'(n)$ в различные моменты времени можно определить результирующее среднее искажение:

$$D = \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{i=1}^L d[\vec{S}_i, \vec{S}'_i], \quad (6.18)$$

где L – общее число векторов.

Если векторный процесс $\vec{S}(n)$ обладает свойствами стационарности и эргодичности, выборочное среднее (6.18) в пределе стремится к математическому ожиданию (6.19):

$$\begin{aligned} D &= M[d(\vec{S}, \vec{S}')] = \sum_{i=1}^L P(\vec{S} \in C_i) M[d(\vec{S}, \vec{S}'_i) | \vec{S} \in C_i] = \\ &= \sum_{i=1}^L P(\vec{S} \in C_i) \int_{s \in C_i} d(\vec{S}, \vec{S}'_i) p(\vec{S}) ds, \end{aligned} \quad (6.19)$$

где $P(\vec{S} \in C_i)$ – вероятность того, что $\vec{S}$ находится в C_i ; L – количество векторов отображений $\vec{S}'_i$; для входного вектора $\vec{S}_i$, $p(\vec{S})$ – многомерная функция плотности вероятности.

При передаче каждый вектор $\vec{S}'_i$ кодируется двоичным кодовым словом W_i , содержащим B_i бит. В общем случае различные кодовые слова имеют различную длину. При этом скорость передачи V определяется выражением:

$$V = BN_k \text{ бит/с}, \quad (6.20)$$

где $B = \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{n=1}^L B(n)$ – бит/вектор,

то есть B определяет среднюю длину кодового слова, $B(n)$ – число бит, использованных для кодирования вектора $\vec{S}_i$, N_k – число кодовых слов, передаваемых за одну секунду.

При использовании КК размером L максимальное число бит для кодирования каждого входного вектора, определяется выражением:

$$B_{\max} = \log_2 L. \quad (6.21)$$

Модель ВК позволяет учесть четыре свойства компонент вектора, которые, будучи использованы должным образом при составлении КК, обеспечивают оптимальные рабочие характеристики. Вот эти свойства: линейная зависимость, нелинейная зависимость, форма ФПВ и многомерность. Все четыре свойства в некоторой степени взаимосвязаны.

Для заданного размера КК указанные выше четыре свойства используются при ВК путем правильного расположения кодовых векторов в N -мерном пространстве, что достигается, например, обучением.

Сжатие данных основывается на исключении избыточности (для передачи избыточной информации требуются дополнительные биты). Избыточность обычно предполагает наличие некоторой зависимости между передаваемыми параметрами. Статистические зависимости можно разделить на два типа: линейная зависимость и нелинейная зависимость. Линейная зависимость представляет собой то, что обычно называют корреляцией. Две коррелированные случайные переменные линейно зависимы. Если две переменные некоррелированы, то они не являются линейно зависимыми, хотя статистическая зависимость может иметь место. Эту «остаточную» зависимость (т. е. то, что остается после удаления линейной зависимости) называют нелинейной зависимостью. Две переменные (с нулевым средним) S_1 и S_2 не коррелированы, если математическое ожидание их скалярного произведения равно нулю [77]:

$$M\{|S_1| \times |S_2|\} = 0. \quad (6.22)$$

Но S_1 и S_2 независимы, когда их совместная ФПВ равна произведению частных плотностей переменных S_1 и S_2 [95]:

$$p(S_1, S_2) = p(S_1)p(S_2), \forall (S_1, S_2). \quad (6.23)$$

Если S_1 и S_2 некоррелированы, но зависимы, то такую зависимость называют нелинейной.

В алгоритме ВК основным принципом принятия решения является поиск вектора-эталона по правилу ближайшего соседа. Данный выбор осуществляется либо путем полного перебора векторов в КК, либо с помощью эффективных быстрых методов поиска [114]. Полный перебор позволяет получить минимальные искажения, но требует большого объема вычислений и памяти.

При построении системы сжатия данных квантователь должен проектироваться таким образом, чтобы для заданной скорости передачи искажения выходного сигнала были минимальны [19]. Процесс построения КК часто реализуется через процесс обучения или заполнения КК. Для построения КК с L кластерами N -мерное пространство разделяется на L ячеек $\{C_i, 1 \leq i \leq L\}$ и каждой ячейке C_i приписывается вектор $\vec{S}$. Тогда в квантователе назначается кодовый вектор $\vec{S}'_i$, если $\vec{S}$ лежит в ячейке C_i . Квантователь называется оптимальным (обеспечивающим минимальные искажения), если мера искажения минимизирована по всем L -уровням квантователя. Оптимальный ВК определяется минимизацией меры искажения по всем L – уровням квантования. При этом существуют два условия оптимальности. Первое заключается в применении правила выбора по минимуму искажений, т.е. выбора ближайшей ячейки

$$q(S_i) = S'_i \text{ тогда, когда } 1 \leq i \leq L. \quad (6.24)$$

При равенстве искажений на разных векторах назначается некоторое правило, позволяющее разрешить эту неопределенность и осуществить выбор.

Второе условие состоит в том, что каждый кодовый вектор $\vec{S}'_i$ выбирается из условия минимизации среднего искажения в ячейке C_i .

$$D_i = \varepsilon \left[d(\vec{S}_i, \vec{S}'_i) / \vec{S}_i \in C_i \right] = \int_{\vec{S}_i \in C_i} d(\vec{S}_i, \vec{S}'_i) p(\vec{S}_i) d(\vec{S}_i), \quad (6.25)$$

а вектор $\vec{S}'_i$ называется центроидом ячейки C_i – $\vec{S}' = \text{cent}(C_i)$. Практически для СКО или ВСКО среднее искажение D_i минимизируется при

$$\vec{S}'_i = \frac{1}{M} \sum_{\vec{S} \in C_i} \{\vec{S}\}, \quad (6.26)$$

где $\{\vec{S}\}$ – набор обучающих векторов, а $\vec{S}'_i$ представляет собой выборочное среднее обучающих векторов, содержащихся в ячейке C_i .

Для расчета центроида необходимо задать меру искажений. Ячейки, определяемые таким образом, известны как ячейки ближайших соседей или области Дирхле [176].

На практике задается набор обучающих векторов $\{\{\vec{S}(i)\}, 1 \leq i \leq Q\}$, где Q – количество обучающих векторов. Некоторое подмножество N_i этих векторов попадает в ячейку C_i , тогда среднее искажения D_i определяется по следующей формуле:

$$D_i = \frac{1}{N_i} \sum_{\vec{S} \in C_i} d(\vec{S}, \vec{S}'_i), \quad (6.27)$$

где N_i – количество векторов попавших в ячейку C_i .

Существует несколько алгоритмов построения КК, однако наиболее часто применяемым в приложениях телекоммуникаций является алгоритм К-средних.

Алгоритм К-средних. Построение КК использует итерационный кластерный алгоритм, известный в литературе по распознаванию образов как алгоритм К-средних. В нашем случае $K = L$, где L – размер КК. Алгоритм разделяет набор обучающих векторов $\{\vec{S}(i)\}$ на L кластеров C_i . Ниже через m обозначен номер итерации, а через $C_i(m)$ – i -й кластер на m -й итерации с центроидом $\vec{S}'_i(m)$. Алгоритм построения КК К-средних записывается в следующем виде.

Шаг 1. Задание начальных значений. Положим $m=0$. Выберем тем или иным подходящим методом набор начальных кодовых векторов $\vec{S}'_i(0)$, $1 \leq i \leq L$.

Шаг 2. Классификация. Классифицируем набор обучающих векторов $\{\{\vec{S}(i)\}, 1 \leq i \leq Q\}$, по кластерам C_i с помощью правила ближайшего соседа $\vec{S} \in C_i(m)$ тогда и только тогда, когда $d[\vec{S}, \vec{S}'_i(m)] < d[\vec{S}, \vec{S}'_j(m)] \quad \forall i \neq j$.

Шаг 3. Коррекция кодового вектора. $m \leftarrow m+1$. Произвести коррекцию кодовых векторов всех кластеров путем вычисления центроидов обучающих векторов каждого кластера, как показано в (6.28):

$$\vec{S}'_i(m) = \text{cent}(C_i(m)), \quad 1 \leq i \leq L. \quad (6.28)$$

Шаг 4. Проверка на окончание процедуры. Если уменьшение величины общего искажения. $d(m)$ на итерации m относительно $d(m-1)$ меньше некоторого порога ΔD_0 , процедура заканчивается. Иначе переход на шаг 2.

На шаге 4 можно использовать какую-либо другую подходящую проверку на окончание процедуры [114].

Путем ряда приближений, задавая кодовым векторам различные начальные значения и повторяя алгоритм для нескольких наборов начальных значений, с последующим выбором КК для обеспечения минимума полных искажений можно достигнуть глобального оптимума. Следовательно, корректное формирование КК (обучение, заполнение КК) обеспечивает устойчивую работу ВК при заданном минимуме искажений входного вектора.

После построения КК ее можно использовать для квантования каждого входного вектора. Вычислительная сложность алгоритма ВК зависит от сложности поиска вектора-эталона по правилу ближайшего соседа. Использование метода полного перебора повышает вычислительную сложность алгоритма нахождения вектора-кандидата, в связи, с чем применяются методы иерархической кластеризации, например методы дихотомии [114]. В ходе исследований алгоритмов построения КК и работы ВК РС был разработан новый способ реструктуризации КК, понижающий вычислительную сложность процедуры ВК [135]. Процедура ВК является

неотъемлемой частью современных алгоритмов кодирования РС. В представленных технических решениях она используется для кодирования параметров ЛП и отдельных сегментов обрабатываемого РС. Современные алгоритмы обработки РС включают процедуру ВК как способ понижения скорости передачи при сохранении приемлемых показателей качества синтезированной речи [215]. В предлагаемых методах обработки ВК используется для кодирования параметров элементов декомпозиции РС $(\vec{w}, \vec{e})$ при их кластерном представлении в CELP подобных алгоритмах.

При квантовании $\vec{S}$ в $\vec{S}'$ возникает ошибка квантования, определяемая мерой искажения $d(\vec{S}, \vec{S}')$ (мера расхождения, мера расстояния). При обработке РС получили распространение следующие меры:

1. Среднеквадратическая ошибка (СКО)

$$d_2(\vec{S}, \vec{S}') = \frac{1}{N} (\vec{S} - \vec{S}')^T (\vec{S} - \vec{S}') = \frac{1}{N} \sum_{k=1}^N (S_k - S'_k)^2, \quad (6.29)$$

где искажение определяется по каждой координате.

2. Общая мера искажений

$$d_r(\vec{S}, \vec{S}') = \frac{1}{N} \sum_{k=1}^N |x_k - y_k|^r \quad (6.30)$$

при $r = 1$ и $r = \infty$ представляет собой среднее значение абсолютной ошибки и максимальной ошибки.

3. ВСКО

$$d_w(\vec{S}, \vec{S}') = (\vec{S} - \vec{S}')^T W (\vec{S} - \vec{S}'), \quad (6.31)$$

где W - положительно определенная взвешивающая матрица, увеличивающая вклад искажений некоторых из параметров в общую меру искажений, данная мера известна также под расстоянием Махаланобиса [114].

4. Мера Итакуры-Саито

Эта мера основана на принципе максимального правдоподобия и рассчитывается по формуле

$$d(\vec{S}, \vec{S}') = (\{\vec{S}\} - \{\vec{S}'\})^T \{B_s\} (\vec{S} - \vec{S}'), \quad (6.32)$$

где $\{B_s\} = \{B(i-k)/B(0), 0 \leq i, k \leq N-1\}$ – нормированная корреляционная матрица, коэффициенты $B(i-k)$ которой являются значениями автокорреляционной функции, используемой для расчета вектора коэффициентов предсказания.

Коэффициенты корреляции нормированы по $B(0)$, соответственно матрица B_s и вектор значений $\vec{S}$ определяют друг друга. B_s выступает как весовая матрица, при этом значения B_s изменяются при изменении S . Так как B_s и $B_{s'}$ различны, то данная мера является несимметричной, т.е. $d(S, S') \neq d(S', S)$.

ВК может обеспечить значительно лучшие характеристики, чем скалярное квантование, особенно для РС $(\vec{w}, \vec{e})$, имеющего нелинейные зависимости. Для процедуры ВК важными характеристиками выступают вычислительная сложность и емкость устройств памяти. При L -уровневом ВК число операций необходимое для квантования одного входного вектора равно L . Вычислительные затраты на квантование каждого входного вектора и емкость памяти для хранения кодовых векторов определяются

$$\xi = NL = N2^{\log_2 L}, \quad (6.33)$$

где N – число операций умножения сложения для каждого расчета искажений.

Затраты на вычислительную сложность и емкость памяти могут быть снижены при использовании некоторых принципов построения и поиска в КК (таблица 6.3) [19, 135].

Таблица 6.3 – Способы построения и поиска в кодовой книге при ВК

Способ	Вычислительные затраты	Емкость памяти
ВК с древовидным поиском (дихотомия)	$\xi = 2NB$	$\mu = 2N(L-2)$
Каскадное ВК	$\xi = N(L_1 + L_2)$	$\mu = N(L_1 + L_2)$
Мультипликативное ВК	$\xi = (N_1 + N_2)L_1L_2$	$\mu = N_1L_1 + N_2L_2$
Случайное ВК	$\xi = NL = N2^{\log_2 L}$	$\mu = NL = N2^{\log_2 L}$

Наиболее качественные характеристики по критерию СКО обеспечивает ВК с полным перебором (рисунок 6.8), но существенные затраты на вычислительную сложность и емкость памяти вынуждают использовать алгоритмы быстрого поиска при некоторых ухудшениях характеристик. Анализ сокращения затрат на вычислительную сложность и емкость запоминающих устройств (таблица 6.3, рис. 6.8) показывает потенциальные возможности использования каскадного ВК и его разновидностей с применением новых технологий построения КК и поиска в них [34, 49, 140].

Процедура ВК в технике кодирования РС связана, как правило, с преобразованием двух параметров модели обработки РС: вектора сигнала возбуждения и вектора параметров ЛП (или иных коэффициентов формирующей модели). При передаче коэффициентов, рассчитанных на сегменте квазистационарности, происходит передача соответствующего данному сегменту и коэффициентам непосредственно самого сигнала возбуждения, либо информации о нем в виде индексов КК, рассчитанных при помощи процедуры «анализа через синтез».

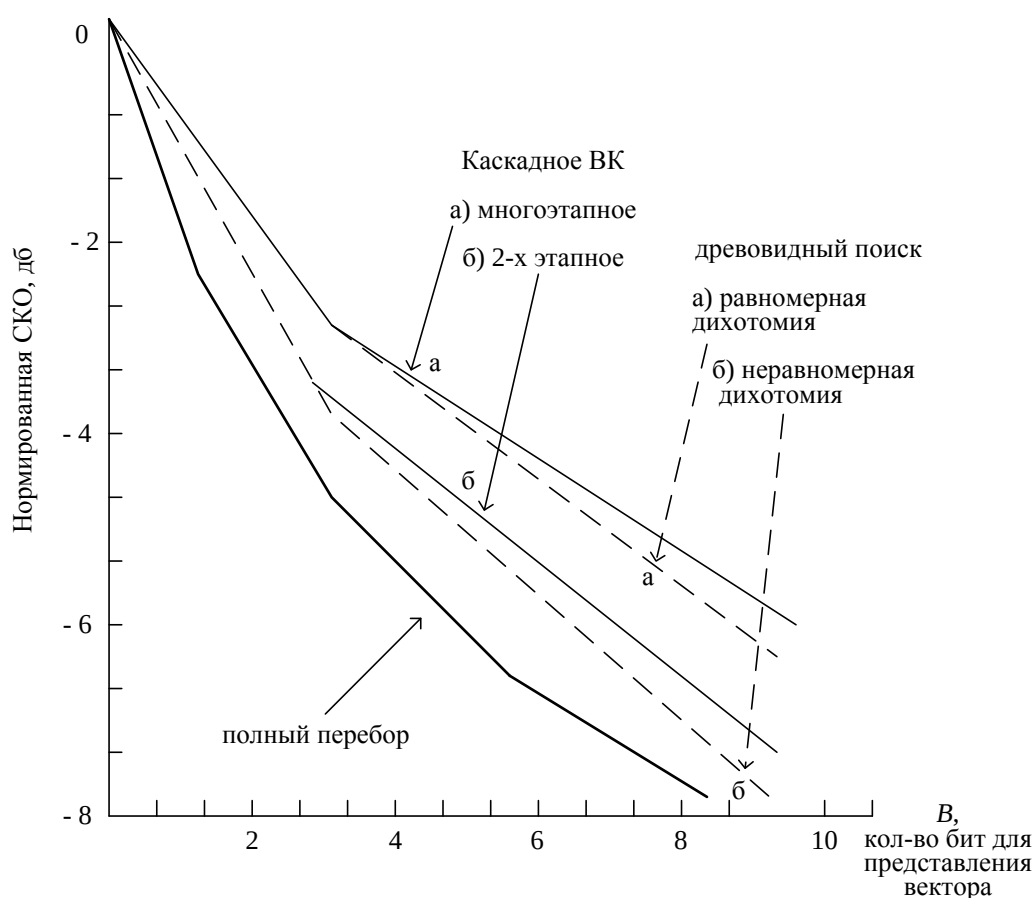


Рисунок 6.8 – Сравнение СКО квантования параметров формирующей модели при линейном предсказании ($N=10$)

Анализ метода ЛП РС и его модификаций показывает, что в современных стандартах кодирования РС:

- не учитываются взаимные связи между сигналом возбуждения фильтра синтеза и вектором параметров передаточной функции голосового тракта, при этом значительное число бит в кадре передачи отводится для представления сигнала возбуждения (таблица 3.1);

- в рассмотренных алгоритмах кодирования не используется априорная информация о характере сигнала, т. е. отсутствует классификация сегментов анализа по различным признакам, использование которой в предлагаемых технических решениях дает возможность совершенствования комплексов обработки РС в перспективных системах связи.

6.5.2 Алгоритм установления взаимосвязей элементов декомпозиции речевого сигнала

При цифровой обработке РС эффективное кодирование речевых данных для их передачи по цифровым КС является одной из основополагающих задач. Ее решение позволяет увеличить пропускную способность линейных трактов и каналов передачи при заданных критериях качества связи. Одним из принципов построения систем обработки и передачи речи выступает снижение скорости передачи при сохранении качественных показателей синтезированного РС. Как было отмечено в разделе 3 при реализации процедуры предсказания и использования в качестве пространств представлений элементов декомпозиции ВК формируются КК фиксированной мощности. Между соответствующими векторами данных пространств возникают зависимости, использование которых позволяет по-новому рассмотреть задачу обработки РС на основе метода ЛП [10, 14, 47, 51, 53, 58].

При обработке РС исключается процедура поиска сигнала возбуждения синтезирующего фильтра на основе классической процедуры "анализа через синтез", которая реализуется путем оценки всех хранящихся в КК векторов сигналов возбуждения.

Теоретические и экспериментальные исследования ЛП речи показали, что ясно прослеживаются зависимости между остатком ЛП и параметрами, описывающими передаточную функцию голосового тракта, которые являются параметрами модели формирования речи в вокодере с ЛП. Данные связи между вектором, содержащим параметры голосового тракта, и соответствующим вектором сигнала ошибки ЛП, рассчитанным на интервале квазистационарности, позволяют избавиться от необходимости передачи по КС сигналов возбуждения, являющихся входным сигналом фильтра синтеза РС на приемной стороне. Данные зависимости можно использовать для понижения скорости передачи информационного потока по КС, применяя на стороне передачи и приема специальным образом обученные КК, предназначенные для решения задачи идентификации соответствующего сигнала возбуждения.

При анализе сегмента РС на передающей стороне создается вектор параметров описывающих передаточную функцию голосового тракта и вектор сигнала ошибки ЛП. В ходе процесса классификации устанавливается однозначная взаимосвязь между кластерами векторов возбуждения ЛП и кластерами векторов параметров голосового тракта для каждого сегмента РС [14]. По КС передаются только параметры модели обработки, коэффициент усиления и данные, характеризующие кодируемый РС. Для формирования сигнала возбуждения в вокодере на основе ЛП на приеме из кадра передачи выделяют параметры для синтезирующего фильтра, содержащие информацию о параметрах голосового тракта, данных РС, а также значение коэффициента усиления сигнала возбуждения. По данным параметрам выделяется соответствующий им класс сигналов возбуждения. По нему идентифицируют подпространство векторов сигнала возбуждения в котором находится вектор кандидат, который используется в синтезирующем фильтре вокодера с ЛП для формирования сегмента РС на однородном участке стационарности. Блок-схема алгоритма нахождения зависимостей элементов декомпозиции речи при ЛП представлена на (рисунке 6.9). Извлеченный сигнал возбуждения используют в синтезирующем фильтре липредера для формирования цифрового РС на сегменте анализа.



Рис.6.9 – Блок-схема алгоритма нахождения зависимостей элементов декомпозиции речи при линейном предсказании

Для этого на первом этапе задаются мощности пространств представлений КК сигналов возбуждения и параметров, описывающих передаточную функцию го-

лосового тракта и размерности соответствующих векторов. Обучение (формирование) КК производится на основании классического алгоритма К-средних. Далее осуществляется выделение зависимостей между соответствующими векторами пространств представлений элементов декомпозиции РС. Данная процедура реализуется в ходе работы с тестовым материалом. Выделяются области КК сигнала возбуждения, которые в процессе процедуры анализа через синтез выбираются как наилучшие для обработки РС при уникальном векторе КК параметров вектора голосового тракта [47,48].

В результате данной процедуры формируются подпространства зависимых представлений векторов параметров голосового тракта и соответствующего им сигналов возбуждения. Далее формируется массив данных, в котором хранятся многозначные не обратные отображения $\varphi: A \rightarrow E$, являющиеся основой для реализации процедур поиска вектора сигнала возбуждения соответствующего вектору параметров голосового тракта из подпространства КК векторов сигналов возбуждения. При использовании предлагаемого подхода кодовые книги, хранящие вектора сигналов возбуждения дообучаются, при этом выявляются подпространства кодовой книги, которые участвуют в формировании синтезированного речевого сигнала и связаны с исходным вектором параметров голосового тракта, таким образом, внесение искажений за счет использования данного подхода минимально.

Таким образом, существует объективная возможность значительного уменьшения мощности подпространства представления сигналов возбуждения, что дает возможность значительно снизить скорость передачи данных в КС, а также уменьшить вычислительную сложность алгоритма кодирования РС на передающей стороне [15]. Использование предлагаемого метода направлено на снижение скорости передачи в вокодерах с ЛП и позволяет понизить скорость передачи данных в КС от известных решений, либо перераспределить информационный ресурс, предоставляемый КС, на формирование дополнительных сервисов абонентского обслуживания. При этом качественные показатели синтезированной речи не претерпевают значительного ухудшения по отношению к исходному методу на основе

CELP, так как меняется не сам способ кодирования, а за счет реструктуризации кодовых книг реализуется изменение представления элементов декомпозиции на основе учета их зависимости.

6.5.3 Кодирование (декодирование) речевого сигнала с учетом зависимости элементов декомпозиции речевого сигнала

На основании предложенной в разделе 3 модели системы обработки РС, учитывающей взаимосвязи элементов его декомпозиции при обработке на основе метода ЛП разработаны новые способы кодирования РС [139, 143]. Схема системы обработки (кодирования) речи и использованием двухуровневой классификации представлена на (рисунке 6.10).

РС, ограниченный верхней частотой 3,4 кГц, подвергается кодированию с использованием линейной ИКМ с $f_d=8$ кГц.

Использование такого диапазона частот РС на входе линейной ИКМ обосновывается необходимостью выделения параметров СТШ и частоты ОТ, диапазон изменения которой представлен в подразделе 6.2.

1. Формируется сегмент анализа длительностью 20 мс, что соответствует фрагменту речи, состоящему из 160 отсчетов РС.

2. Производится начальная обработка речевого сегмента анализа (фильтрация). Сегмент центрируется посредством вычитания из всех отсчетов РС математического ожидания при условии равновероятности всех отсчетов в сегменте [77].

3. В соответствии с блок-схемой алгоритма, представленной на рисунке 4.12, производится классификация речевых сегментов на конечное число классов [2, 6, 45]. Каждый сегмент активной речи приписывается к одному из 7 типов сегментов. Принцип классификации сегментов рассмотрен в 6.3.

4. После классификации каждый сегмент подвергается процедуре идентификации фильтра синтеза ЛП РС, заключающейся в расчете параметров формирующей модели. Данная процедура подробно рассмотрена в [40,44,55,57].

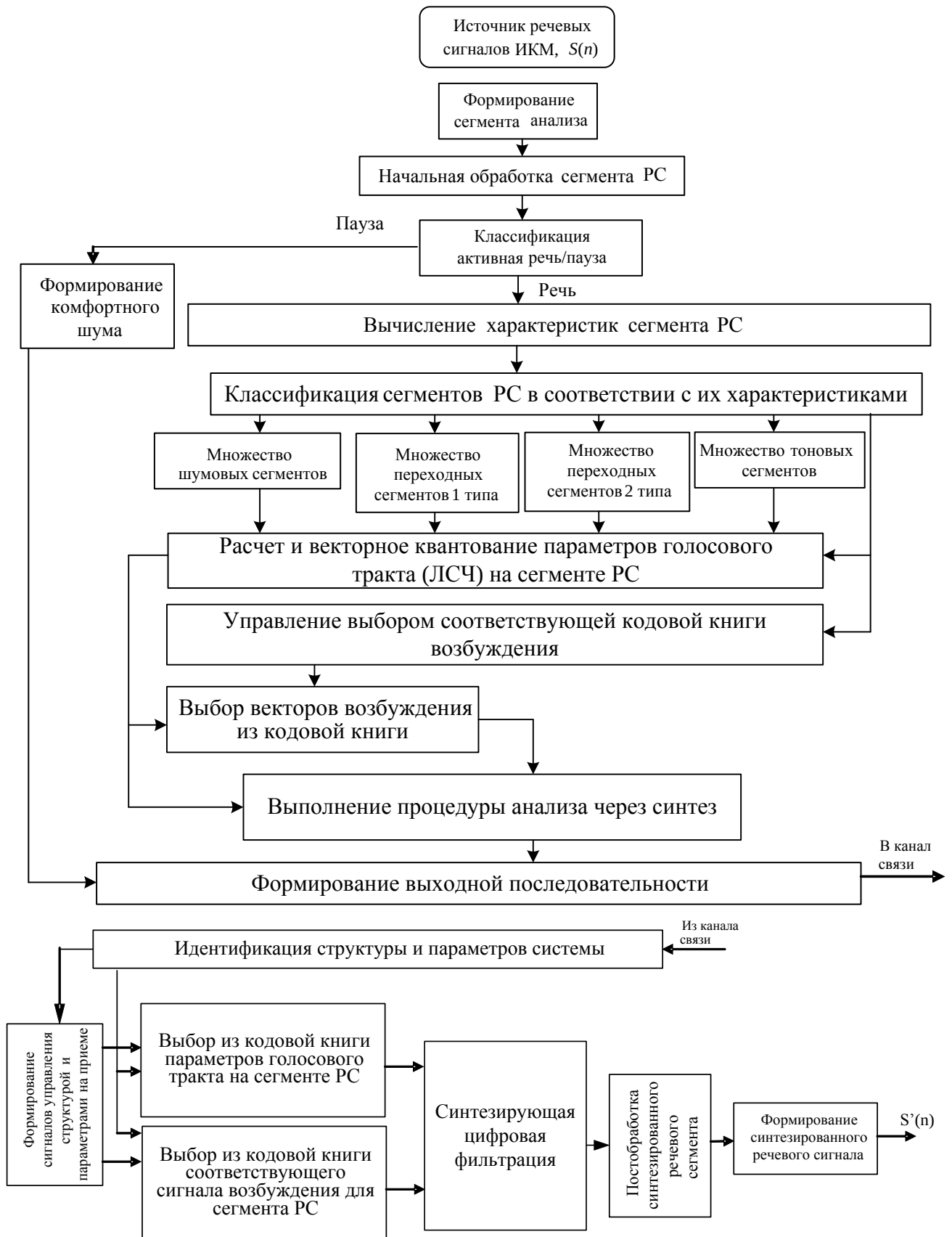


Рисунок 6.10 – Система кодирования (декодирования) РС учитывающая зависимости элементов декомпозиции при его обработке на основе линейного предсказания

5.1. В случае принятия решения "пауза" система обработки формирует комфортный шум.

5.2. Параметры формирующей модели и соответствующий сигнал возбуждения для тоновых, переходных 1 и 2-го типа, шумовых сегментов обрабатываются с использованием процедур ВК. Для каждого вида сегмента в соответствии с рассмотренной выше классификацией используются свои КК ВК. Процедуры функционирования ВК, настройки и алгоритмы обучения КК ВК рассмотрены в подразделе 6.5.1. На этапе функционирования применяются заранее обученные КК ВК с конечным числом классов разделений пространств параметров декомпозиции сегментов РС. Этап обучения КК ВК параметров формирующей модели РС и сигналов возбуждения позволяет определить зависимость между данными параметрами декомпозиции. Таким образом, между пространством классификации (кластеров) параметров формирующей модели и пространством векторов сигнала ошибки ЛП (сигнала возбуждения формирующей модели) существует взаимосвязь, которую можно использовать для снижения скорости передачи [9, 15, 51, 52, 53, 58]. Пример установления взаимосвязи для элементов декомпозиции РС при их кластерном представлении на основе ЛП и их дальнейшем ВК представлен в (таблице 6.4).

Таблица 6.4 – Соответствие векторов элементов декомпозиции РС

Номера векторов кодовой книги параметров голосового тракта	Номера векторов кодовой книги сигналов возбуждения										
1	18	25	34	38	54	69	125	203	250	—	...
2	36	42	50	54	56	67	92	—	—	—	...
3	71	76	80	84	89	93	144	189	226	245	...
4	5	18	104	200	212	—	—	—	—	—	...
5	9	32	40	—	—	—	—	—	—	—	...
...	...			...		...		...	...	...	...

Следует отметить, что при установлении взаимосвязей определить точное соответствие вектора параметров голосового тракта и вектора сигнала возбуждения

точно для всех векторов не представляется возможным, что связано с особенностями обучения соответствующих КК. Исследования сформированных КК на тестовом материале, показало, что максимальное число векторов возбуждения для одного вектора ЛСЧ может достигать 16-20, для кодовой книги состоящей из 256 векторов сигналов возбуждения. (рисунок 6.11) [10, 42]. Таким образом, сокращается количество бит для представления вектора в подпространстве КК сигналов возбуждения, что дает возможность до 20% уменьшать среднюю скорость передачи в КС при сохранении, либо незначительном ухудшении, качественных показателей синтезированного РС в процессе низкоскоростного кодирования.

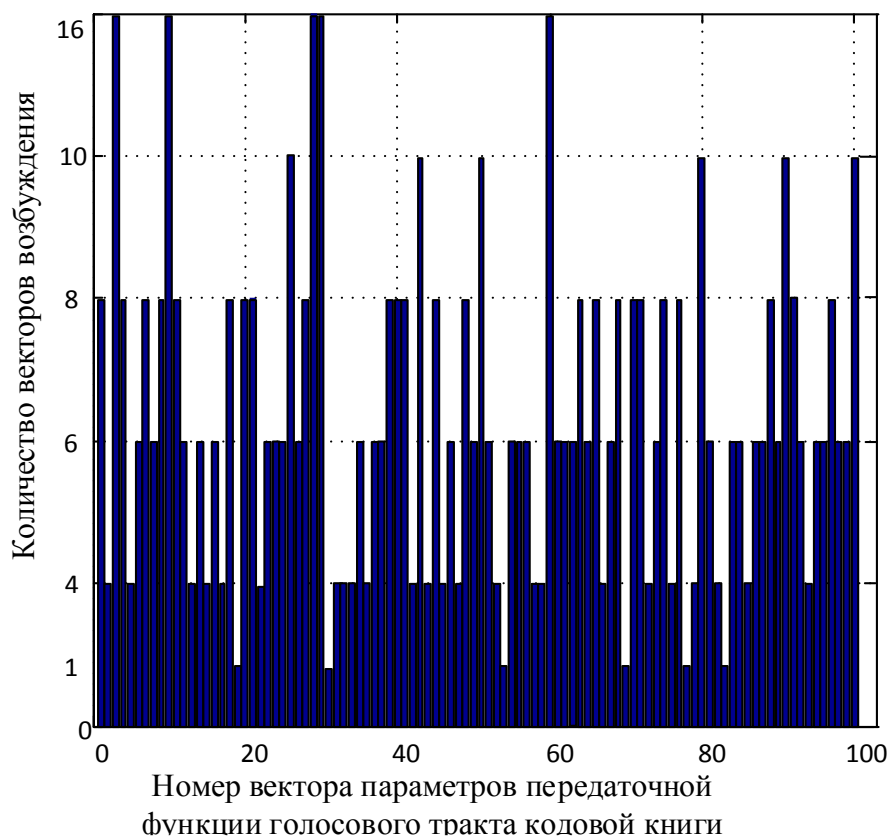


Рисунок 6.11 – Распределение векторов возбуждения для вектора ЛСЧ при процедуре «анализа через синтез»

6. Производится формирование кадра передачи, максимальная скорость при таком подходе может составить 2,4 кБит/с и определяется структурой кадра передачи. Для достижения данной скорости необходимо выделить и распределить максимально 48 бит для формирования отдельного сегмента РС, так как число анализируемых в секунду сегментов равно 50.

7. На приеме по принятой комбинации производится выбор структур необходимых КК ВК, соответствующих классам разделения на тоновые, переходные и шумовые сегменты. На основе процедуры поиска в кластеризованных пространствах параметров декомпозиции из соответствующих КК извлекается информация о векторах параметров модели обработки и возбуждения.

8. Производится обработка сегмента РС с помощью выбранных элементов декомпозиции сегмента РС. Процедуры обработки сегмента РС для ЛП описаны во втором разделе. На стыках сегментов при их обработке используется интерполяция отсчетов. Данная процедура предназначена для устранения паразитной частоты векторной дискретизации [114].

9. Сигнал подвергается дополнительной обработке (постфильтрация) [188].

10. Цифровой сигнал подается на цифроаналоговый преобразователь линейной ИКМ для формирования выходного восстановленного аналогового РС.

Блок-схема алгоритма функционирования системы обработки (кодирования) РС представлена на (рисунке 6.12) (передающая часть), (рисунке 6.13) (приемная часть).

Для дальнейшего уменьшения скорости передачи возможно либо уменьшение числа бит, предоставляемых на описание информационных параметров, либо увеличение длительности сегмента анализа до 20 мс (160 отсчетов) для рассмотренного выше кадра передачи.

При переходе от одного сегмента к другому возникает необходимость сохранения начальных условий фильтра синтеза, так как память о предыдущем сегменте в алгоритме ЛП является основой для формирования текущего отсчета на выходе синтезирующей системы. Обновление памяти фильтра синтеза в модели ЛП происходит с использованием последних отсчетов предыдущего сегмента. Начальные условия фильтра синтеза после сегмента «пауза» принимаются соответствующими комфортному шуму [188] системы обработки.

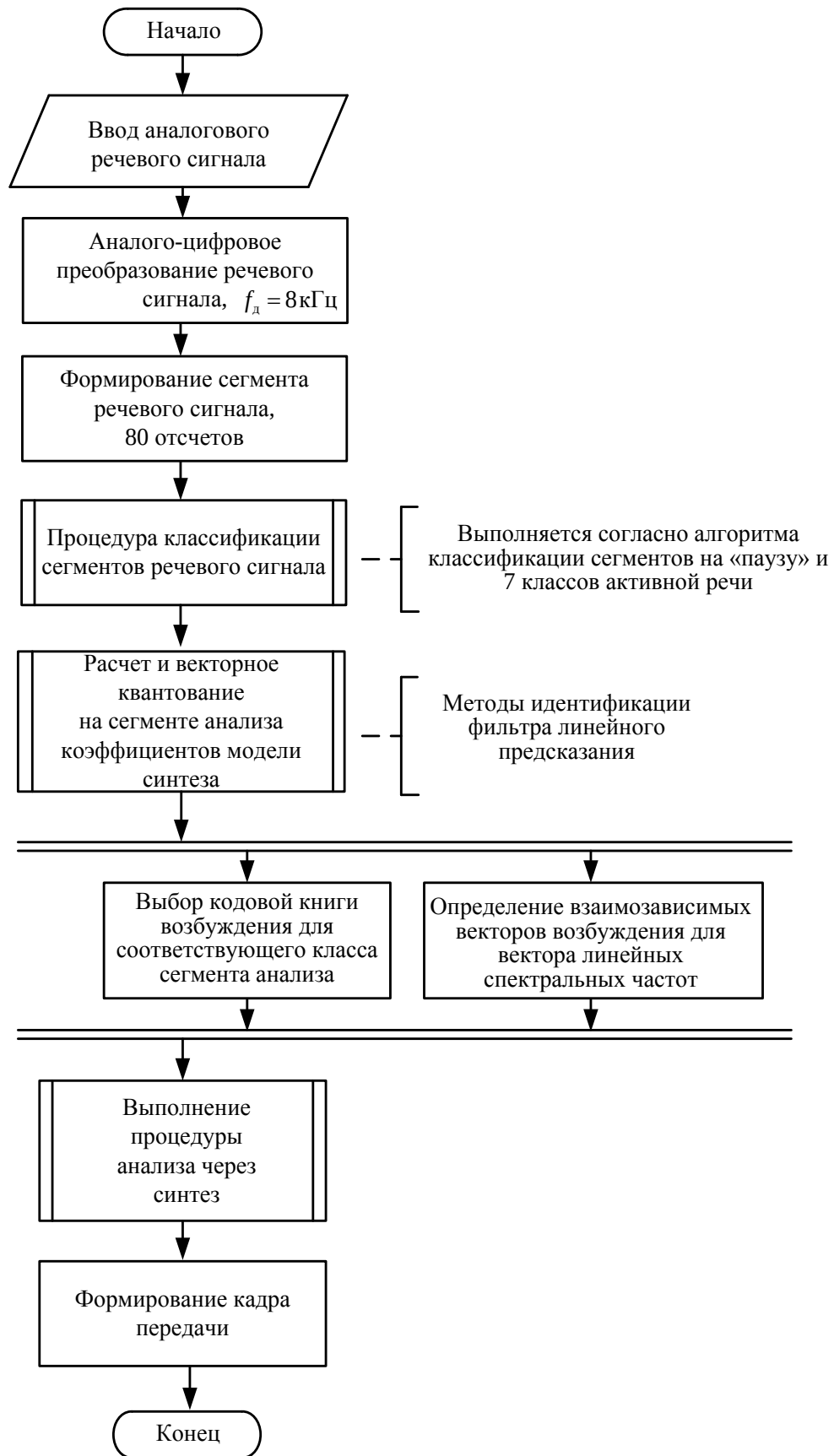


Рисунок 6.12 – Блок-схема алгоритма функционирования передающей части системы кодирования РС



Рисунок 6.13 – Блок-схема алгоритма функционирования приемной части системы кодирования РС

Таким образом, на основе анализа свойств РС с учетом зависимостей элементов декомпозиции РС представлена система обработки РС, отличающаяся от известных решений тем, что в ней реализуются процедуры перестройки параметров и структуры системы с использованием процедур ВК элементов декомпозиции сегментов РС.

Одной из задач, решаемых при разработке алгоритмов функционирования системы обработки сигналов, является проверка их корректности в соответствии с предъявляемыми к ним требованиями [173].

Основными требованиями к алгоритмам являются:

1. Алгоритм должен содержать конечное количество элементарно выполнимых операций.
2. Алгоритм должен выполнять конечное количество шагов при решении задачи, т.е. удовлетворять требованию конечности действий.
3. Алгоритм должен быть единым для всех допустимых исходных данных, т.е. удовлетворять требованию универсальности.
4. Алгоритм должен приводить к корректному по отношению к поставленной задаче решению, т.е. удовлетворять требованию правильности.

Тестирование разработанных алгоритмов выполнено на реальных РС длительностью не менее 15 минут для 13 дикторов. Все операции, представленные в алгоритмах, описаны как простейшие (элементарные действия), каждое из которых разделяется на элементарные действия, описание которых представлено в [188].

При подобном техническом решении снижение скорости передачи реализуется за счет уменьшения мощностей подпространств представления КК сигналов возбуждения. Учет данной зависимости при построении системы обработки РС дает возможность существенно уменьшить мощность векторного подпространства представления сигналов возбуждения синтезирующего фильтра, связанного с вектором параметров голосового тракта. Мощность векторного подпространства представления сигналов возбуждения связанных с вектором параметров голосового тракта и определяет в итоге малую размерность КК векторов сигналов возбуждения. Сравнение известных технических решений с предлагаемыми показывает следующее:

- 1) В обоих случаях выполняется кодирование речевого сигнала известным методом линейного предсказания с использованием векторного квантования.
- 2) В обоих случаях выполняется передача номеров векторов из кодовых книг, отражающих сигнал возбуждения и параметры голосового тракта.

Основное отличие предлагаемого технического решения от известных способов кодирования РС заключается в использовании выявленной взаимной статистической связи между вектором параметров голосового тракта и сигналом возбуждения, что позволяет снизить размерность сигнального пространства векторов сигнала возбуждения для его представления. Такое решение дает возможность снизить количество бит, выделяемых для векторного квантования сигналов возбуждения, а, соответственно, снизить скорость передачи в КС в среднем на 20% (таблица 5.10). При выполнении процедуры *AbS* поиск в КК соответствующего вектора проводится не по всей КК, а по ее подпространству, что дополнительно снижает требование к вычислительной сложности алгоритмов обработки РС. К недостаткам данного подхода следует отнести увеличение требований к объему запоминающих устройств для хранения таблиц соответствия векторов элементов декомпозиции РС между собой.

6.6 Особенности реализации предложенных методов обработки речевого сигнала в системах телекоммуникаций

Анализ качества разработанных алгоритмов обработки РС с учетом зависимостей элементов декомпозиции произведен на основе оценки параметров разработанной системы. Необходимость перестройки структуры системы кодирования, обусловила введение в комплекс оцениваемых параметров емкости устройств памяти, необходимой для реализации в речевом кодеке процедуры кластеризации и классификации элементов декомпозиции РС. Очевидно, что введение процедуры классификации сегментов РС привело к увеличению вычислительной сложности разработанных алгоритмов [46].

Значения временной задержки на кодирование РС разработанными алгоритмами являются ориентировочными оценками, рассчитанными на основе вычислительной сложности алгоритмов с учетом их структуры и пиковой производительности современных семейств ЦПОС и СнК, рекомендуемых в качестве элементной

базы для реализации систем обработки РС. При расчете задержки кодирования должны учитываться следующие компоненты задержки: задержка АЦП/ЦАП, задержка формирования сегмента анализа РС, алгоритмическая задержка. Точные величины τ при заданных значениях вычислительной сложности зависят от конкретного варианта реализации гибридного кодека РС: архитектуры процессора, степени ее ориентированности на операции ЦОС, реальной производительности процессора, качества отладки программного продукта при его трансляции в машинный код и т.д. Однако рассчитанные ориентировочные значения τ позволяют утверждать о непревышении в разработанных алгоритмах существующих значений алгоритмической задержки для стандартов кодирования. Переход к классификации сегментов РС наряду с ВК элементов декомпозиции требует увеличения объема хранимой в устройствах памяти информации, что обусловлено необходимостью хранения параметров КК. Увеличение требуемого объема устройств памяти представляет собой основной вид затрат и одновременно является основным (но не критичным с точки зрения реализации) недостатком разработанных алгоритмов по сравнению с существующими алгоритмами кодирования РС. Совершенствование процесса гибридного кодирования РС процедуры классификации ВК коэффициентов формирующей модели и сигнала возбуждения, обеспечивающей перестройку структуры системы кодирования, позволяет разрабатывать РПУ, соответствующие традиционным задачам разработки речевых кодеков для инфокоммуникаций. В основе улучшения кодека РС разработанными алгоритмами лежит углубленный анализ сегмента РС, характеризующийся возрастанием вычислительной сложности процедур кодирования речи. На аппаратном уровне разработанные алгоритмы по сравнению с существующими алгоритмами требуют повышенной производительности вычислителей и дополнительной емкости элементов памяти [30].

Использование и комбинация аппаратных и программных средств реализации систем с ЦОС, к классу которых принадлежат предлагаемая система, определяют аппаратный, программный и аппаратно-программный способы реализации вычислительных алгоритмов [172].

Решение о выборе конкретного класса сверхбольших интегральных схем (СБИС) для реализации разработанных алгоритмов кодирования не может быть

принято без учета некоторых особенностей, к числу которых следует отнести:

- необходимость обработки РС в режиме реального времени;
- высокую скорость поступления исходных данных;
- широкий динамический диапазон входных и выходных сигналов;
- большое количество операций сложения, умножения и логических операций;
- параллелизм вычислительных процедур на некоторых этапах обработки РС;
- повторяемость многих вычислительных операций на длительности сегмента анализа РС;
- применение достаточно сложного математического аппарата и технологий обработки речи.

Указанный перечень специфических особенностей предлагаемых алгоритмов позволяет сделать вывод о возможности использования ЦПОС для реализации системы обработки РС [32]. Современный рынок микропроцессорных устройств имеет в наличии большое количество разнообразных процессоров, существенно отличающихся друг от друга по характеристикам. ЦПОС используются в самых различных областях, предъявляющих к процессорам совершенно разные требования. Естественно, не существует идеального процессора для всех областей применения. Существенно важными характеристиками для выбора того или иного процессора для реализации разработанных алгоритмов являются:

1) Технология производства. Современные ЦПОС, как правило, изготавливаются по КМОП-технологии (комплементарная структура "металл-окисел-полупроводник") с технологической нормой 0,25-0,09 мкм. Следует иметь в виду, что технология производства ЦПОС характеризует степень интеграции СБИС и во многом обуславливает их основные характеристики, а тенденция уменьшения технологической нормы устойчиво сохраняется.

2) Формы представления чисел. Различные процессоры используют различные формы представления чисел: либо с фиксированной точкой, либо с плавающей точкой; отдельные семейства ЦПОС поддерживают обе формы. Очевидно, что

форма с ПТ является более гибкой и удобной при разработке большого числа систем ЦОС, однако соответствующие процессоры являются более сложными и дорогими устройствами.

3) Форматы представления чисел. Большинство популярных процессоров используют кодовые слова длиной в 16, 24, 32, 40 и 64 бита. Кроме того, при осуществлении вычислительных процедур активно используются форматы двойного слова и различные варианты расширенного слова.

4) Код представления чисел. Подавляющее большинство ЦПОС, ориентированных на форму с ФТ, используют дополнительный код, а в процессорах, ориентированных на форму с ПТ, как правило, применяется прямой код.

5) Внутренняя тактовая частота процессора во многом влияет на производительность ЦПОС. Очевидно, что период тактовой частоты однозначно определяет длительность командного цикла процессора (времени выполнения простейшей команды типа «регистр, регистр → регистр», означающей, что источником двух операндов при ее выполнении являются регистры и результат операции также помещается в регистр). Так, длительность командного цикла процессора, внутренняя тактовая частота которого составляет 200МГц, равна 5 нс. Следует отметить, что некоторые семейства ЦПОС имеют встроенные системы деления или умножения частоты, при этом внутренняя тактовая частота работы процессора не совпадает с внешней, подаваемой на ЦПОС. Тактовая частота и длительность командного цикла недостаточно характеризуют производительность процессора. Дело в том, что, с одной стороны, отдельные операции в процессоре могут выполняться за несколько циклов, а с другой – в некоторых процессорах используется параллельное выполнение команд и параллельная работа нескольких независимых операционных модулей.

6) Производительность ЦПОС характеризует скорость обработки данных в процессоре и имеет пиковое и реальное значения.

Пиковая производительность представляет собой теоретический максимум быстродействия процессора при идеальных условиях, который определяется как число вычисли-

тельных команд, выполняемых в единицу времени всеми имеющимися в процессоре арифметическими устройствами. Предельное быстродействие достигается при обработке бесконечной последовательности команд, не связанных между собой и не конфликтующих при доступе в память. Разумеется, подобная ситуация является чисто гипотетической, на практике ни один процессор не в состоянии работать сколько-нибудь длительное время с пиковой производительностью. Наиболее популярными показателями пиковой производительности являются количество миллионов команд (инструкций), выполняемых процессором за секунду (MIPS – Million of Instructions Per Second), при представлении данных в форме с ФТ и количество миллионов операций, выполняемых за секунду при арифметике с ПТ (MFLOPS – Million of Floating-point Operations Per Second). В простейшем случае количество MIPS (MFLOPS) совпадает со значением тактовой частоты, выраженной в мегагерцах. Иногда пиковая производительность ЦПОС измеряется в миллионах операций в секунду (MOPS) или в миллионах операций умножения с накоплением в секунду (MMAC). Повышение пиковой производительности ЦПОС относительно тактовой частоты достигается архитектурными решениями, направленными на увеличение степени параллелизма в обработке данных, что проявляется в возможности одновременного выполнения в процессоре нескольких команд или операций. К таким решениям, в первую очередь, относятся:

- увеличение количества функциональных модулей процессора (АЛУ, умножителей);
- увеличение количества дополнительных специализированных устройств и сопроцессоров (например, переход к многоядерным процессорам);
- увеличение разрядности шин передачи данных и программ для увеличения количества одновременно передаваемой информации.

Применение технологий полиспектрального анализа при реализации процедур функционирования предлагаемой системы обработки в условиях наличия акустических шумов определяет повышенные требования к вычислительной производительности используемых аппаратных средств.

Перспективы разработки и внедрения отечественной радиоэлектронной аппара-

туры на основе государственной программы импортозамещения на основе серий отечественных интегральных микросхем открываются благодаря разработкам, ведущимся сегодня в ГУП НПЦ «Элвис» и НТЦ «Модуль» [239, 240]. Среди базовых серий микросхем разработки НПЦ «ЭЛВИС» следует отметить серию программируемых сигнальных контроллеров «Мультикор». Микросхемы данной серии (таблица 6.5) – это однокристалльные программируемые многопроцессорные «системы на кристалле» (СНК) на базе платформы «Мультикор».

Таблица 6.5

Сравнение микросхем серии «Мультикор» с зарубежными ЦПОС

Фирма	ЭЛВИС			TI		ADI	
Процессор	МС-24	МС-0226	МСF-0428	С6701	С6416	TS201	BF53
Тактовая частота, МГц	100	120	400	166	600	600	600
Пиковая производительность, форматы: • 16b, int (MOPs) • 32b, float (MFLOPs) • 8b, int (MOPs)	1600 600 3600	3840 1440 8640	25600 9600 57600	– 1000 –	4800 – –	9600 3600 –	3360 – –
Время выполнения КИХ-фильтрации 35-го порядка, 1024 входных отсчета (мкс): • int (MAC: 16*16+32) • int (MAC: 8*8+32) • 32 b, float • 32b, float, 15 отводов, 1024 входных отсчетов	99.8 50.4 195.4 90.1	41.6 21 81.4 37.6	6.3 3.2 12.3 5.6	– – 393 –	25.7 – – –	– – 32.8 18.6	– – – –
Время расчета FFT-1024, компл. (мкс): • 16b, block float. • 32b, float.	58 107	24.2 44.6	3.7 6.7	– 160	11 –	– 19.3	– –
Время расчета FFT-256, компл. (мкс): • 16b block float • 32b, float	11.2 21.5	4.7 9.0	0.7 1.4	– –	– –	0.975 4.12	5.3 –
Декодер Витерби, 16b, int (MOPs)	200	480	3200	–	–	–	–
Время расчета SIN(X)/COS(X), 32b, float (мкс): вектор, 128 отсчетов	0.09	0.04	0.006	–	–	0.02	–
Время расчета квадратного корня, 32b, float (мкс)	0.08	0.03	0.005	–	–	0.007	–

СБИС НПЦ «ЭЛВИС» принадлежат к классу СнК, что обусловлено наличием на кристалле как центрального процессорного устройства CPU, так и ЦПОС. CPU представляет собой *Reduced Instruction Set Computer (RISC)*-ядро, является ведущим и имеет доступ к ресурсам ЦПОС. ЦПОС, входящие в состав СнК, имеют гарвардскую архитектуру, функционируют под управлением CPU и предназначены для эффективной реализации алгоритмов ЦОС.

Важным преимуществом микросхем серии «Мультикор» является характеристика плотности кодов для DSP-ядер, которая, как правило, лучше на порядок, чем у зарубежных аналогов [239, 240]. Команда для DSP-ядра достаточно емкая, так как имеет длину 32 или 64 бита, и напоминает с этой точки зрения архитектуру VLIW (Very Long Instruction Word). К примеру, одна 64-разрядная инструкция для ELcore-24™ может выполнять две пары арифметических (логических) операций любой сложности и две пары пересылок типа регистр-память. Система инструкций DSP обеспечивает программирование всех базовых процедур сигнальной обработки. DSP-ядро ELcore-xx™ функционирует под управлением RISC-ядра и расширяет его возможности по обработке сигналов. Система команд DSP-ядра обеспечивает программирование всех базовых процедур сигнальной обработки. DSP-ядро имеет типичную для ЦПОС гарвардскую архитектуру с внутренним параллелизмом по потокам обрабатываемых команд и данных. Все процессоры обеспечены общим полем памяти на регистровом уровне или на уровне внутренней памяти и работают независимо друг от друга (каждый по своей собственной программе), и, вследствие этого, представляют СнК MIMD-архитектуры (MIMD – Multiple Instructions Multiple Data).

Для повышения производительности ядра ELcore-xx™ используется распараллеливание потоков обработки по SIMD-типу (SIMD – Single Instructions Multiple Data). Это достигается благодаря масштабируемости DSP-ядра ELcore-xx™, то есть возможности увеличения числа секций обработки данных (SIMD-секций) от одной (ELcore-1x™) до двух (ELcore-2x™), оставаясь в рамках одной системы инструкций и одной среды разработки. Микросхемы серии «Мультикор» являются импортозамещающими и способны конкурировать с аналогичными зарубежными образцами (таблица 6.6).

Таким образом, можно сделать вывод о соответствии отечественных микропроцессоров платформ «Мультикор» уровню реализации алгоритмов обработки РС в РПУ системы связи, которые в настоящее время используют микропроцессоры TMS320 [153, 238].

Таблица 6.6 – Характеристики некоторых СБИС, производимых ГУП НПЦ ЭЛВИС

Процессор	Технологическая норма, мкм	Центральный процессор (RISC-ядро)		Вычислительные ядра ЦОС (ЦПОС)				
		тип ядра	тактовая частота, МГц	тип ядра	формы чисел	форматы чисел	тактовая частота, МГц	MIPS (MFLOPS)
1892BM3T (MC-12)	0,25	RISCore 32	80	Одно ядро EL-core-14 архитектуры SISD	ФТ	8	100	1 440
						16		640
						32		320
					ПТ	32		240
1892BM2Я (MC-24)	0,25	RISCore 32	80	Одно ядро ELcore-24 архитектуры SIMD	ФТ	8	100	2 880
						16		1 280
						32		640
					ПТ	32		480
1892BM5Я (MC-0226)	0,25	RISCore 32	100	Два ядра ELcore-26 архитектуры SIMD	ФТ	8	120	7 200
						16		3 200
						32		1 600
					ПТ	32		1 200
1892BM7Я* (MCF-0428)	0,18	RISC-ядро	250	Четыре ядра архитектуры SISD	ФТ	8	250	33 800
						16		24 000
					ПТ	32		6 000
MC-24R	0,25	RISC-ядро	100	Одно ядро архитектуры SIMD	ФТ	8	100	3 600
						16		1 600
						32		800
					ПТ	32		600
MCom-01	0,13	RISCore 32	333	Четыре ядра ELcore-28	ФТ	8	333	47 800
						16		32 000
MCom-02	0,13	RISCore 32	140	Два ядра ELcore-17	ФТ	8	280	10 080
						16		4 480
						32		2 240
					ПТ	32		1 680
NVCom-01	0,13	RISC-ядро	300	Два ядра архитектуры SISD	ФТ	8	300	24 000
						16		14 400
					ПТ	24		3 600
						32		1 200

Есть основания предполагать, что реализация предложенных в работе алгоритмов обработки РС на СБИС платформ «Мультикор» обеспечит создание высоконадежных систем обработки РС перспективных отечественных комплексов системы связи, превосходящих существующие системы их лучшие зарубежные аналоги. Жизненный цикл таких систем будет измеряться десятилетиями за счет возможности программной и аппаратной модернизации (путем перепрограммирования либо за счет создания новых проблемно-ориентированных СнК на базе программно совместимых с предыдущими поколениями ИМС отечественного производства).

Можно выделить ряд направлений в которых могут использоваться перспективные технические решения в области обработки РС в интересах абонентов систем связи:

- модернизация сети подвижной радиосвязи;
- повышение качества обслуживания мобильных абонентов за счет внедрения современного абонентского, оконечного и коммутационного оборудования;
- организация удаленной обработки данных мобильного абонентского терминала в системах передачи с использованием методов коммутации каналов и коммутации пакетов;
- разработка и внедрение новых технических средств, использующих методы пакетной коммутации, предоставление абонентам широкого спектра услуг связи с одного абонентского устройства;
- создание сегмента сети связи с внедрением оконечного оборудования нового поколения, поддерживающего функции абонентского шифрования и обработки РС в специальных условиях функционирования;
- разработка мультимедийных мобильных абонентских средств;
- дальнейшее совершенствование федеральной подсистемы сотовой связи стандарта GSM;
- внедрение адаптивного комплекса передачи речи и данных по КВ каналам, включающего подвижные и стационарные средства КВ связи различной мощности;

– создание сегмента сети связи для предоставления абонентам мультисервисных услуг информационного обеспечения в различных условиях функционирования.

Таким образом, в существующих и разрабатываемых образцах перспективных комплексов связи присутствует необходимость в обработке РС и его передаче в диапазоне скоростей 0,6-9,6 кбит/с.

В рамках изложенных предложений по применению разработанных технических решений обработки РС наиболее рациональным с точки зрения технической реализации будет использование элементной базы российского производства.

Представленные технические решения по обработке РС делают адекватным их применение в рамках существующих и перспективных сетевых технологий передачи. До недавнего времени сетевые технологии не обеспечивали непосредственного обслуживания РС, а известный метод предварительной буферизации с целью выравнивания скорости генерирования цифрового потока не только значительно уменьшал преимущества рассматриваемых кодеков, но и в условиях низкоскоростного кодирования речи создавал ряд дополнительных трудноразрешимых сложностей. Современное состояние инфокоммуникационных систем и тенденции их развития свидетельствуют о признании права на существование и перспективности представленных технических решений по модернизации и совершенствованию устройств обработки РС.

6.7 Выводы по шестому разделу

1. Наличие в гибридных алгоритмах кодирования РС процедуры ВК дает возможность выявить зависимости элементов декомпозиции РС и позволяет разработать алгоритмические решения для задачи совершенствования системы обработки РС в условиях воздействия акустических шумов различной природы, по качественным характеристикам превосходящие решения, используемые в существующих средствах инфокоммуникаций.

2. Разработанные технические решения по обработке РС по сравнению с существующими, требуют повышенной производительности вычислителей и дополнительной емкости элементов памяти.

3. Реализация разработанных технических решений по обработке РС на СБИС российского производства серий «Мультикор обеспечит создание высоконадежных перспективных мобильных комплексов связи.

4. Представленные технические решения свидетельствуют о перспективности применения новых процедур в технике обработки РС при использовании зависимостей элементов декомпозиции РС, переменной длины сегмента анализа в условиях воздействия шумовых акустических воздействий.

5. Проверка основных свойств алгоритмов фильтрации речевого сигнала в условиях интенсивных акустических воздействий свидетельствует о возможности их практической реализации.

6. Представленные технические решения свидетельствуют о перспективности применения разработанных моделей и методов обработки РС при наличии акустических шумов с целью улучшения функционирования УОРС в целом.

ЗАКЛЮЧЕНИЕ

В диссертационной работе сформулирована научно-техническая проблема, решение которой дало возможность развить теоретические принципы анализа РС, на основании которых предложить модели и методы совершенствования систем обработки РС, учитывающие особенности природы его возникновения и принятые технологии обработки в системах инфокоммуникаций.

Представленные технические решения подтверждают выдвинутую гипотезу исследования о том, что использование положений теории линейной фильтрации и результатов полиспектрального анализа РС в условиях наличия акустических шумовых воздействий, а также учет зависимости элементов его декомпозиции и особенностей речеобразования при его моделировании в системах связи, дают возможность улучшить характеристики функционирования систем его обработки.

Основные результаты выполненной работы представлены ниже.

1. Показана актуальность модернизации существующих систем обработки РС, выявлены объективные предпосылки снижения средней скорости передачи в КС без ухудшения качества синтезированной речи.

2. Представлено комплексное техническое решение по формированию методологии понижения скорости кодирования РС при его передаче с сохранением качественных показателей синтезированной речи в условиях наличия акустических шумовых воздействий;

3. Развита теоретическая модель обработки РС как модели в виде свертки параметров, описывающих передаточную функцию ограниченного порядка голосового тракта и сигнала возбуждения, которые отличаются от известных методов учетом зависимости элементов декомпозиции РС, динамическим изменением длительности сегмента анализа при коррекции значений параметров синтезирующего фильтра голосового тракта.

4. Разработаны:

- модель кодирования РС со структурно-параметрической адаптацией, основанная на его представлении в виде свертки параметров, описывающих передаточную функцию ограниченного порядка голосового тракта и сигнала возбуждения,

отличающаяся от известных учетом особенностей формирования речевого сигнала и технологий его кодирования, и позволяющая разработать методологию снижения средней скорости передачи данных в КС.

- модель обработки речевого сигнала, основанная на учете зависимости элементов декомпозиции речевого сигнала при его низкоскоростном кодировании на основе метода линейного предсказания, позволяющая осуществить снижение мощности пространства представления сигналов возбуждения.

- модель системы фильтрации РС, отличающаяся применением полиспектрального анализа и позволяющая разработать методы шумоподавления в речевом сигнале в широком диапазоне значений ОСШ.

5. Предложены:

- метод выделения сегментов анализа речевого сигнала различной длительности на однородных участках речевого сигнала, позволяющий осуществлять синтез речевого сигнала при фиксированных значениях параметров передаточной функции голосового тракта;

- метод обработки речевого сигнала, основанный на учете зависимости элементов декомпозиции РС при линейном предсказании, позволяющий уменьшать среднюю скорость передачи в канале связи при сохранении качественных показателей синтезированной речи.

- метод адаптивной цифровой фильтрации речевого сигнала в условиях акустического зашумления, отличающийся применением технологий полиспектрального анализа и позволяющий осуществлять качественное шумоподавление при низких отношениях сигнал-шум до $0 \div -5$ дБ.

6. Разработаны и представлены алгоритмы анализа и обработки РС, позволяющие повысить характеристики функционирования телекоммуникационных устройств в системах инфокоммуникаций в условиях наличия акустических шумов, отличающиеся от известных учетом особенностей формирования и обработки РС, применением результатов полиспектрального анализа с целью шумоподавления, а также учетом зависимости элементов декомпозиции РС при линейном предсказании, устраняющие противоречия между особенностями речеобразования и существующими в современных системах связи технологиями обработки.

Востребованность предложенных в диссертации элементов теории, моделей, методов и алгоритмов анализа и обработки РС подтверждается результатами решения исследовательских задач, направленных на практическое применение предлагаемых технологий обработки РС в интересах инфокоммуникационной среды Российской Федерации.

Полученные в диссертационной работе результаты соответствуют пунктам "2. Исследование процессов генерации, представления, передачи, хранения и отображения аналоговой, цифровой, видео-, аудио- и мультимедиа информации; разработка рекомендаций по совершенствованию и созданию новых соответствующих алгоритмов и процедур", "8. Исследование и разработка новых сигналов, модемов, кодеков, мультиплексоров и селекторов, обеспечивающих высокую надежность обмена информацией в условиях воздействия внешних и внутренних помех", "11. Разработка научно-технических основ технологии создания сетей, систем и устройств телекоммуникаций и обеспечения их эффективного функционирования" и "12. Разработка методов эффективного использования сетей, систем и устройств телекоммуникаций в различных отраслях народного хозяйства" паспорта специальности 05.12.13 – Системы, сети и устройства телекоммуникаций.

Значение диссертации для экономического развития и обороноспособности страны определяется тем, что в ее рамках исследованы и показаны методы улучшения качественных показателей функционирования устройств обработки речевого сигнала, применение которых определяет пути совершенствования инфокоммуникационных систем государства, как в мирное время, так и в условиях чрезвычайных ситуаций, вооруженных конфликтов. В диссертационной работе раскрыты перспективные направления развития устройств обработки РС в инфокоммуникационных системах, применение которых даёт возможность перейти на новый этап использования подобных технологий обработки РС. На основании полученных результатов работы можно сделать вывод о достижении поставленной цели диссертационного исследования.

СПИСОК СОКРАЩЕНИЙ

АЦП	– аналого-цифровое преобразование
ВК	– векторное квантование
КВ	– коротковолновой
КК	– кодовая книга
КРС	– кодек речевых сигналов
КЛП	– коэффициент линейного предсказания
КС	– канал связи
КСБ	– коэффициент суммарной бикорреляции
ЛП–	– линейное предсказание
ЛСК	– линейный спектральный корень
ЛСЧ	– линейная спектральная частота
МПАТ	– мобильный персональный абонентский терминал
МСЭ	– международный союз электросвязи
ОСШ	– отношение сигнал/шум
ОТ	– основной тон
РПУ	– речепреобразующее устройство
РС	– речевой сигнал
СБИС	– сверхбольшие интегральные схемы
СКО	– среднеквадратическое отклонение
СнК	– система на кристалле
СПМ	– спектральная плотность мощности
УИАВ	– условия интенсивных акустических воздействий
УОРС	– устройство обработки речевого сигнала
ФПВ	– функция плотности вероятности
ЦАП	– цифроаналоговое преобразование
ЦК	– цифровой канал
ЦОС	– цифровая обработка сигналов
ЦПОС	– цифровой процессор обработки сигналов
AbS	– <i>Analysis-buy-Synthesis</i>

<i>ACELP</i>	– <i>Algebraic Code Excited Linear Prediction</i>
<i>AMR</i>	– <i>Adaptive Multi-Rate</i>
<i>AMR-WB</i>	– <i>Adaptive Multi-Rate Wideband</i>
<i>CD</i>	– <i>Cepstral distance</i>
<i>CELP</i>	– <i>Code Excited Linear Prediction</i>
<i>CS-ACELP</i>	– <i>Conjugate-Structure and Algebraic CELP</i>
<i>GSM</i>	– <i>Global System for Mobile telecommunications</i>
<i>IP</i>	– <i>Internet Protocol</i>
<i>ISD</i>	– <i>Itakura-Saito Distance</i>
<i>LD-CELP</i>	– <i>Low Delay CELP</i>
<i>LPC</i>	– <i>Linear Predictive Coding</i>
<i>RPE-LTP-</i>	– <i>Rich Pulse Excitation Long Term Prediction</i>
<i>MBSD</i>	– <i>Modified Bark Spectral Distortion</i>
<i>MFLOPS</i>	– <i>Million of Floating-point Operations Per Second</i>
<i>MIPS</i>	– <i>Million of Instructions Per Second</i>
<i>MOS</i>	– <i>Mean Opinion Score</i>
<i>MPE-LPC</i>	– <i>Multi Pulse Excitation Linear Predictive Coding</i>
<i>MPLS</i>	– <i>Multiprotocol Label Switching</i>
<i>RELP</i>	– <i>Residual Excited Linear Prediction</i>
<i>RISC</i>	– <i>Reduced Instruction Set Computer</i>
<i>SSD</i>	– <i>Segmental Spectral Distortion</i>
<i>VAD</i>	– <i>Voice Activity Detector</i>

СПИСОК ЛИТЕРАТУРЫ

1. Абрамов, В. А. Информация и звук : учеб. пособие / В. А. Абрамов. МТУСИ. – М., 2008. – 93 с.
2. Айвазян, С. А. Прикладная статистика : классификация и снижение размерности/С. А. Айвазян, В. М. Бухштабер.–М.:Финансы и статистика, 1989.–608с.
3. Алдошина, И. А. Электроакустика и звуковое вещание: учеб. пособие для вузов/И.А.Алдошина [и др.].–М.:Горячая линия-Телеком,2007.–872 с.
4. Афанасьев, А.А. Выделение сегментов квазистационарности при анализе речевого сигнала в вокодерах с линейным предсказанием/ А.А. Афанасьев, В.Г. Трубицын // Фундаментальные и прикладные проблемы техники и технологии. – 2010. – № 2(280). – С. 97–103.
5. Афанасьев, А.А. Синтез сигналов возбуждения в низкоскоростных вокодерах с линейным предсказанием/А.А. Афанасьев//Телекоммуникации. –2010. – №11. – С.8–13.
6. Афанасьев, А.А. Психоакустический классификатор кадров при низкоскоростном кодировании широкополосного речевого сигнала /А.А. Афанасьев, О.О.Басов, М.В. Илюшин // Фундаментальные и прикладные проблемы техники и технологии. – 2010. – № 3(281). – С. 73–78.
7. Афанасьев, А.А. Использование психоакустической модели слуха при разработке вокодеров с линейным предсказанием/ А.А. Афанасьев, М.В. Илюшин // Цифровая обработка сигналов.– 2011.– № 1. – С. 49–52.
8. Афанасьев, А.А. Уменьшение объема данных при широкополосном кодировании речевого сигнала/А.А. Афанасьев, В.Г. Трубицын// Цифровая обработка сигналов. – 2011. – № 1. – С. 53–56.
9. Афанасьев, А.А. Метод снижения скорости передачи в вокодерах с линейным предсказанием на основе применения нейронных сетей при переменной длине сегмента анализа / А.А. Афанасьев, А.П. Рыжков // Информационные системы и технологии. – 2012. – № 6 (74). – С. 20 – 28.

10. Афанасьев, А.А. Использование взаимозависимостей параметров линейного предсказания при реализации процедур обработки речевых данных/ А.А. Афанасьев, А.П. Рыжков //Телекоммуникации.– 2012.– № 13. – С. 32–36.

11. Афанасьев, А.А. Модель системы предобработки речевого сигнала на основе полиспектрального анализа / Афанасьев А.А., Титов О.Н. // Вестник Рязанского государственного радиотехнического университета– 2014– Вып.50.– С. 9–16.

12. Афанасьев, А.А. Непрерывная аутентификация диктора при ведении телефонных переговоров по низкоскоростным цифровым каналам / А.А. Афанасьев // Вопросы кибербезопасности.– 2015. – № 3(16).– С. 60–68.

13. Афанасьев, А.А. Непрерывная аутентификация легитимного пользователя сети связи, опирающаяся на использование личностных особенностей низкоскоростного кодирования речевых данных/ А.А. Афанасьев// Труды научно-техн. конф. кластера пензенских предприятий, обеспечивающих безопасность информационных технологий. Пенза-2016.– Том 10. – С. 66–73.

14. Афанасьев, А.А. Перспективы развития систем низкоскоростного кодирования речи при учете взаимозависимостей элементов декомпозиции речевого сигнала/ А.А. Афанасьев // Информационные системы и технологии.–2016.– №2(94).– С.14-21.

15. Афанасьев, А.А. Система обработки речевого сигнала на основе модернизации метода прямого предсказания / Афанасьев, А.А. // Проектирование и технология электронных средств. – 2017.– №3.– С.

16. Афанасьев, А.А. Модель системы предобработки речевого сигнала на основе накопления базы биспектров сигналов речи / А.А. Афанасьев, Д.Д. Сергеев, О.Н. Титов // Информационные системы и технологии. – 2018.– №2. – С.56-67.

17. Афанасьев, А. А. Теоретико-множественная модель липредера с адаптацией распределения информационных ресурсов к статистическим характеристикам кодируемых параметров/ А.А. Афанасьев, Г.В. Богачев, А.А. Рыболовлев// Деп. в ВИНТИ, Академия ФАПСи. 2002. – Орел. – 60 С.

18. Афанасьев, А.А. Обработка речевых данных на основе использования переменной длины сегмента анализа / А.А. Афанасьев // Сборник статей 2-й региональной научно-практической конференции "Многоядерные процессоры, параллельное программирование, системы обработки сигналов". – Барнаул: АГУ. – 2012 – С.121–125.

19. Афанасьев, А.А. Влияние структуры кодовых книг на работу векторного квантователя при кодировании речи / А.А. Афанасьев, А.В. Пищальников // Материалы 11 Международной научной конференции «Информатизация и информационная безопасность правоохранительных органов». – Москва: 2002. – С. 361–362.

20. Афанасьев, А.А. Адаптивный кодек речевых сигналов на основе систем с переменной структурой / А.А. Афанасьев // Цифровая обработка сигналов и ее применение: материалы 4-й международной конференции. – Москва: ИПУ РАН, 2002. – С.124–126.

21. Афанасьев, А.А. Модель системы с переменной структурой для кодирования речи / А.А. Афанасьев // Информатизация и информационная безопасность правоохранительных органов: материалы 11 Международной научной конференции. – Москва: 2002. – С. 362–365.

22. Афанасьев, А.А. Система адаптивного кодирования речевого сигнала / А.А. Афанасьев, А.В. Пищальников // Цифровая обработка сигналов и ее применение: материалы 5-й международной конференции и выставки. – Москва: ИПУ РАН, 2003. – С. 326–329.

23. Афанасьев, А.А. Помехоустойчивость системы с переменной структурой для кодирования речевых сигналов / А.А. Афанасьев, Г.В. Богачев, Д.Г. Богачев // Современная радиоэлектроника в ретроспективе идей В.А.Котельникова: материалы международной научной конференции к 95-летию академика В.А.Котельникова. – Тезисы докладов. – Москва: Изд. МЭИ, 2003. – С.14-15.

24. Афанасьев, А.А. Особенности системы с переменной структурой для кодирования речевых сигналов / А.А. Афанасьев, Г.В. Богачев, О.О. Басов // Цифровая обработка сигналов и ее применение: материалы 6-й Международной конференции. – Москва: ИПУ РАН, 2004. – С.76–79.

25. Афанасьев, А.А. Анализ возможности перехода к системам с переменной структурой при эффективном кодировании речевого сигнала / А.А. Афанасьев, Г.В. Богачев, М.В. Илюшин // Цифровая обработка сигналов и ее применение: материалы 7-й международной конференции. – Москва: ИПУ РАН, 2005. – С.116–119.

26. Афанасьев, А.А. Применение программы технических расчетов MATLAB для решения задач цифровой обработки сигналов/ А.А. Афанасьев, А.А. Шустров// Материалы 10 Всероссийской научно-техн. конф. студентов, молодых ученых и специалистов. – Рязань: РГРТУ, 2005.– С.44-45

27. Афанасьев, А.А. Построение систем речепреобразования для комплексов телекоммуникаций / А.А. Афанасьев, А.В. Петунин, А.С. Чуйкина // Материалы 11 Всероссийской научно-техн. конф. студентов, молодых ученых и специалистов. – Рязань: РГРТУ, 2006. – С. 62–63.

28. Афанасьев, А.А. Разработка классификатора речевых сегментов / А.А. Афанасьев, А.В. Петунин, А.С. Чуйкина // Материалы 11 Всероссийской научно-техн. конф. студентов, молодых ученых и специалистов.– Рязань, РГРТУ, 2006. – С. 63–64.

29. Афанасьев, А.А. Оптимизация множеств функциональных пространств представлений кодируемых параметров при сжатии данных с потерями / А.А. Афанасьев // Информатизация и информационная безопасность правоохранительных органов: сборник трудов 16-й научной конференции. – Москва: 2007. – С. 370–372.

30. Афанасьев, А.А. Реализация алгоритмов цифровой обработки сигналов на процессорах фирмы Texas Instruments с использованием средств MatLab и LabView // Сети, системы связи и телекоммуникации: материалы 32-й Всероссийской научно-практической конференции. – Рязань: РВВКУС, 2007. – С.142–144.

31. Афанасьев, А.А. Некоторые подходы к построению перспективных систем эффективного кодирования данных с потерями / А.А. Афанасьев, Д.А. Жиров // Материалы 32-й Всероссийской научно-практической конференции "Сети, системы связи и телекоммуникации".–Рязань:РВВКУС,2007.–С.146–148.

32. Афанасьев, А.А. Разработка и отладка систем на основе цифровых модулей Texas Instruments с использованием MATLAB / А.А. Афанасьев, Д.А. Жиров //

Проектирование научных и инженерных приложений в среде MATLAB: труды всероссийской научной конференции.—С-Пб: 2007. — С.1379–1384.

33. Афанасьев, А.А. Использование частотно-временного анализа для представления сигналов возбуждения в вокодерах с линейным предсказанием / А.А. Афанасьев, Д.А. Жиров, И.В. Ульянов // Материалы 33-й Всероссийской научно-практической конференции "Сети, системы связи и телекоммуникации".— Рязань: РВВКУС, 2008. — Часть 1. — С. 135–137.

34. Афанасьев, А.А. Оптимизация координат векторов кодовой книги при обучении / А.А. Афанасьев, И.В. Ульянов // Информтех–2008: материалы 1 всероссийской научно-технической конференции.— Москва:2008.— С. 102–103.

35. Афанасьев, А.А. Синтез сигналов возбуждения в вокодере с линейным предсказанием / А.А. Афанасьев, Д.А. Жиров, И.В. Ульянов // Новые информационные технологии в научных исследованиях — материалы 13 всероссийской научно-технической конференции студентов.— Рязань: РГРТУ, 2008. — Часть 1. — С. 151–153.

36. Афанасьев, А.А. Аппроксимация значений линейных спектральных частот при низкоскоростном кодировании речи /А.А. Афанасьев //Новые информационные технологии в научных исследованиях и в образовании: материалы 15-й Всероссийской научно-технической конференции. — Рязань: РГРТУ, 2010. — С. 88–89.

37. Афанасьев, А.А. Моделирование речевого сигнала на сегментах одинаковой природы образования / А.А. Афанасьев, О.Н. Титов, И.В. Ульянов // Новые информационные технологии в научных исследованиях и в образовании: материалы 16 всероссийской научно-технической конференции студентов, молодых ученых и специалистов. — Рязань: РГРТУ, 2011.— С.77–79.

38. Афанасьев, А.А. Создание интеллектуальных систем индивидуального низкоскоростного речевого кодирования / А.А. Афанасьев // Материалы 8-й Всероссийской научно-технической интернет-конференции. — Тула: Изд. ТулГУ, 2011. — С. 62–63.

39. Афанасьев, А.А. Цифровой спектральный анализ речевого сигнала в приложении к задаче низкоскоростного кодирования / А.А. Афанасьев, О.Н. Титов //

Материалы 8-й Всероссийской научно-технической Интернет-конференции. – Тула: Изд-во ТулГУ, 2011. – С. 63–66.

40. Афанасьев, А.А. Расчет линейных спектральных частот при кодировании речи на сегментах одинаковой природы образования / А.А. Афанасьев // Инновационные технологии: теория, инструменты, практика/ Материалы 3-й международной интернет-конференции (Inno-Tech 2011). – Пермь, 2012. – 5 с. [Электронный ресурс]: [сайт].– Режим доступа: <http://www.altstu.ru/media/f/102.pdf>. – Дата обращения: 5.06.2018.

41. Афанасьев, А.А. Адаптивный байесов подход в задачах обработки речи / А.А. Афанасьев // Информационные технологии. Радиоэлектроника. Телекоммуникации (ITRT-2012) : сборник статей 2-й международной заочной научно-технической конференции. – Тольятти: 2012. – Часть 1. – С. 129–135.

42. Афанасьев, А.А. Влияние конечного порядка АР-модели синтезирующей системы на результаты обработки речевых данных/ А.А. Афанасьев // Современные проблемы радиоэлектроники: всероссийская научно-техн. конф. – Красноярск: 2012. – С. 162–166.

43. Афанасьев, А.А. Формирование сигнала возбуждения в низкоскоростных вокодерах с линейным предсказанием / А.А. Афанасьев, М.В. Илюшин // Цифровая обработка сигналов и ее применение DSPA-2010: материалы 12-й Международной конференции и выставки//Москва: РНТОРЭС, 2010.–С.214-217.

44. Афанасьев, А.А. Выделение параметров модели речевого сигнала на сегментах одинаковой природы образования / А. А. Афанасьев, Б.Р. Иванов, М.В. Илюшин // Цифровая обработка сигналов и ее применение DSPA-2011: Материалы 13-й Международной конференции и выставки.– Т.1, Москва: РНТОРЭС, 2011. – С. 192-194.

45. Афанасьев, А. А. Низкоскоростная передача речевых сигналов с применением нейросетевой классификации в методе линейного предсказания / А. А. Афанасьев, А. П. Рыжков// Проблемы фундаментальных и прикладных естественных и технических наук в современном информационном обществе: сборник трудов 55-й

всероссийской научной конференции МФТИ.– Радиотехника и кибернетика. Том 1. – Москва: МФТИ, 2012. – С. 40-41.

46. Афанасьев, А. А. Реализация нейросетевой обработки речи на современной элементной базе / А. А. Афанасьев, А. П. Рыжков// Молодежь и наука: модернизация и инновационное развитие страны: материалы II международной научно-практической конференции студентов и молодых ученых. – Пенза: ПенГУ, 2012. – С.659–663.

47. Афанасьев, А.А. Выделение зависимостей элементов декомпозиции речевого сигнала при его обработке на основе процедур предсказания / А.А. Афанасьев, А.П. Рыжков // INTERMATIC–2012: материалы 11 Международной научно-технической конференции. –Москва: МГТУ МИРЭА– ИРЭ РАН,2012. – Часть 6. – С. 25–28.

48. Афанасьев, А.А. Устранение априорной неопределенности при обработке речевого сигнала на основе управления мощностями подпространств представления его параметров /А.А. Афанасьев // Материалы докладов конгресса молодых ученых. Вып. 1.– С-Пб: НИУ ИТМО, 2012. – С. 114–116.

49. Афанасьев, А.А. Алгоритмы нейросетевого векторного квантования речевых данных / А. А. Афанасьев, А. П. Рыжков// Инновационные технологии в производстве, науке и образовании: сборник трудов II международной научно-практической конференции. Грозный: Грозненский государственный нефтяной технический университет, 2012. – С.283–292.

50. Афанасьев, А.А. Использование биспектральных оценок для очистки речевого сигнала от шума / А.А. Афанасьев, О.Н. Титов // Современные проблемы радиоэлектроники: материалы докладов 16-й всероссийской с международным участием научно-технической конференции. – Красноярск, 2013. – С. 403–406.

51. Афанасьев, А.А. Уменьшение избыточности представления речевого сигнала при его кодировании на основе метода линейного предсказания/ А. А. Афанасьев // Цифровая обработка сигналов и ее применение DSPA-2013: материалы докладов 15-й международной конференции. Том1. – Москва: ИПУ РАН, 2013. – С. 202–205.

52. Афанасьев, А.А. Локализация подпространства представления сигналов возбуждения синтезирующего фильтра при низкоскоростном кодировании речи / А.А. Афанасьев // Цифровая обработка сигнала и ее применение DSPA-2014: материалы докладов 16-й международной конференции и выставки. Москва: ИПУ РАН, 2014. – Том 1. – С. 191–194.

53. Афанасьев, А.А. Потенциальные возможности понижения скорости передачи в системах кодирования речи / А.А. Афанасьев // Будущее технической науки: сборник материалов XIII международной молодежной научно-технической конференции. – Нижний Новгород, 2014. – С.46–47.

54. Афанасьев, А.А. Частотно-временная сегментация обрабатываемого речевого сигнала на основе полиспектрального анализа /А.А. Афанасьев, О.Н. Титов // Информационные системы и технологии: сборник материалов Международной научно-технической конференции.– Нижний Новгород: НГТУ, 2014.–С. 99–101.

55. Афанасьев, А.А. Вычисление параметров передаточной функции голосового тракта на однородных сегментах речевого сигнала при его низкоскоростном кодировании / А.А. Афанасьев, Р.С. Власов // Системы связи и радионавигации: сборник тезисов III всероссийской научно-технической конференции. – г. Красноярск: АО «НПП «Радиосвязь», 2016. – С.145- 148.

56. Афанасьев, А.А. Классификация сегментов зашумленного речевого сигнала с использованием акустических информационных критериев / А.А. Афанасьев, Титов О.Н.// Современные тенденции и инновации в науке и производстве: сборник материалов 2-й международной научно-практической конференции, г. Междуреченск: КузГТУ, 2014 – С. 245-247.

57. Афанасьев, А.А., Власов, Р.С. Вычисление параметров синтезирующей системы голосового тракта на однородных сегментах анализа при линейном предсказании речевого сигнала / А.А. Афанасьев, Р.С. Власов // Перспективные технологии в средствах передачи информации: материалы 12-ой международной научно-технической конференции. – Владимир: ВЛГУ, 2017.– Том 1.– С. 151-153.

58. Афанасьев, А.А. Снижение размерности кодовой книги сигналов возбуждения в гибридных кодеках с линейным предсказанием на основе учёта зависимости элементов декомпозиции речевого сигнала / А.А. Афанасьев // Перспективные технологии в средствах передачи информации: материалы 12-ой международной научно-технической конференции. – Владимир: ВЛГУ, 2017, том 1. – С.153-156.

59. Афанасьев, А.А. Цифровая обработка сигналов / А.А. Афанасьев, А.А. Рыболовлев, А.П. Рыжков // Учебное пособие. – Москва: Горячая линия - Телеком, 2017. – 356с.

60. Бабкин, В. В. Вокодер 600 бит/с /В. В. Бабкин // Цифровая обработка сигналов и ее применение: материалы 9-й международной конф. – Москва: ИПУ РАН, 2007. – Доклады – IX-1. – С. 157-160.

61. Бабкин, В.В. Шумопонижающее устройство для вокодера /В. В. Бабкин // Цифровая обработка сигналов и ее применение: материалы 9-й международной конф. – Москва: ИПУ РАН, 2007. – Доклады – IX-1. – С. 231-235.

62. Бабков, В. Ю. Передача информации в системах подвижной связи / В. Ю. Бабков, М. А. Вознюк, В. А. Петраков, А. Е. Рыжков, М. А. Сиверс. – С.-Пб. : СПбГУТ, 1999. – 152 с.

63. Баранников, Л. Н. Цифровые методы передачи речевых сообщений / Л. Н. Баранников, А.И. Фомин. – М. : Издательство МАИ, 1996. – 36 с.

64. Беллами, Дж. Цифровая телефония/ Дж. Беллами // Пер. с англ. под ред. А. Н. Берлина, Ю. Н. Чернышова. – М. : Эко – Трендз. 2004. – 640 с.

65. Беллман, Р. Прикладные задачи динамического программирования/ Р. Беллман, С. Дрейфус // Пер. с англ. под ред. А. А. Первозванского. – М.: Наука, гл. ред. физ.-мат. лит-ры, 1965. – 459 с.

66. Богданов, Д. С. База речевых фрагментов русского языка «ISABASE» / Д. С. Богданов, О. Ф. Кривнова, А. Я. Подрабинович, В. В. Фарсобина // Интеллектуальные технологии ввода и вывода информации. – М., 1998.

67. Бокс, Д. Анализ временных рядов / Д. Бокс, Г. Дженкинс// Прогноз и управление – М. : Мир, 1974. – 406 с.

68. Бухалев, В.А. Распознавание, оценивание и управление в системах со случайной скачкообразной структурой. – М. Наука, 1996. – 288с.
69. Быков, С. Ф. Цифровая телефония : учеб. пособие для вузов / С. Ф. Быков, В. И. Журавлев, И. А. Шалимов. – М. : Радио и связь, 2003. – 144 с.
70. Важинский, В.Н. Корреляционная функция третьего порядка и биспектр в задачах обработки сигналов / Важинский В.Н. // ОМП. 1991. – №4. – С. 4-14.
71. Вапник, В.Н. Восстановление зависимостей по эмпирическим данным. М.: Наука. 1979.– 448 с.
72. Вальд, А. Последовательный статистический анализ / А. Вальд // Под редакцией А.Ф. Лапко, Гос.изд. ФИЗМАТЛИТ.– М:-1960. – 328с.
73. Варламова, О. Помехоустойчивые кодеки – будущее цифровой телефонии / О. Варламова // Сети. – 1997. – № 10. – С. 26–32.
74. Величко, В. В. Передача данных в сетях мобильной связи третьего поколения / В. В. Величко // Москва: Горячая линия – Телеком, 2005. – 336 с.
75. Вемян, Г. В. Качество телефонной передачи и его оценка / Г. В. Вемян // Москва: Связь, 1970. – 224 с.
76. Вентцель, Е. С. Теория вероятностей / Е. С. Вентцель // – М. : Высшая школа, 1998. – 576 с.
77. Вентцель, Е. С. Теория вероятностей и ее инженерные приложения / Е. С. Вентцель, Л. А. Овчаров // Учеб. пособие для втузов. – 2-е изд. – Москва: Высшая школа, 2000. – 480 с.
78. Витязев, В. В. Новое в цифровой обработке сигналов / В. В. Витязев // Электросвязь. – 1998. – № 10. – С.15–19.
79. Вокодерная телефония. Под редакцией Пирогова А. А.– Москва: Связь, 1974. – 536 с.
80. Воеводин, В. В. Вычислительные процессы с теплицевыми матрицами / В. В. Воеводин, Е. Е. Тыртышников // – Москва: Наука, 1987. – 168 с.

81. Воробьев, В. И. Спектральные пары в линейном предсказании / В. И. Воробьев, В. Н. Иванов, Д. А. Улахович // Известия ВУЗов. Радиоэлектроника. 1991. – Т. 34. – № 12. – С. 32 – 37.
82. Галунов, В. И. Моделирование процессов передачи информации в звуковом диапазоне / В. И. Галунов, В. Д. Родионов // – Л.: Внешторгиздат, 1988. – 160 с.
83. ГОСТ Р 50840-95. Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости. – Москва: ИПК издательство стандартов, 1996. – 230 с.
84. ГОСТ Р 51061-97. Системы низкоскоростной передачи речи по цифровым каналам. Параметры качества речи и методы измерений. – Москва : Госстандарт России, 1997. – 230 с.
85. Горбатов, В. А. Дискретная математика : учеб. для студентов вузов / В. А. Горбатов, А. В. Горбатов, М. В. Горбатов // Москва: ООО «Издательство АСТ», 2003. – 447 с.
86. Громаков, Ю. А. Стандарты и системы подвижной радиосвязи / Ю. А. Громаков // Москва: МЦНТИ, 1996. – 242 с.
87. Добровидов, А. В. Непараметрическое оценивание сигналов / А. В. Добровидов, Г. М. Кошкин // Москва : Наука. Физматлит, 1997. – 336 с.
88. Дьяконов, В. MATLAB 6 : Учебный курс / В. Дьяконов // СПб.: Питер, 2001. – 592 с.
89. Жиликов, Е. Г. Об одном методе обнаружения пауз в речевых сигналах / Е. Г. Жиликов, С. П. Белов, Е. И. Прохоренко, А. С. Белов / Цифровая обработка сигналов и ее применение: материалы 10-й международной конф. – М. : ИПУ РАН, 2008. – Доклады – X-2. – С. 229–231.
90. Иванов, А. И. Нейросетевой вокодер-архиватор, сохраняющий биометрические особенности голоса говорящего при высоком уровне сжатия шипящих звуков / А. И. Иванов, А. Н. Хальметова, О. С. Захаров, Рыболовлев А. А., Рыжков А. П. // Нейрокомпьютеры: разработка, применение, 2012–№ 3–С.44-48.

91. Иванов, В. Н. Вычисление линейных спектральных частот / В. Н. Иванов // Электросвязь. – 1997. – № 6. – С. 25–27.
92. Иванов, В. Н. Чувствительность спектральных пар / В. Н. Иванов, А. А. Ланнэ, В. Ю. Прокопенко // Известия ВУЗов. Радиоэлектроника. 1991.– Т. 34.– № 12.– С. 37 – 42.
93. Иглин, С. П. Математические расчеты на базе Matlab /С. П. Иглин // – СПб.: BHV-Санкт-Петербург, 2005. – 640 с.
94. Казаков, И.Е. Статистическая динамика систем с переменной структурой / И.Е. Казаков // Москва: Наука, 1977. – 416с.
95. Калинина, В. Н. Математическая статистика : учеб. для студ. сред. спец. учеб. заведений / В. Н. Калинина, В. Ф. Панкин // 4-е изд., испр. – Москва: Дрофа, 2002. – 336 с.
96. Калинин, Ю. К. Разборчивость речи в цифровых вокодерах / Ю. К. Калинин // Москва: Радио и связь, 1991. – 220 с.
97. Кириллов, С. Н. Реализация кодеков речи на основе быстрых ортогональных преобразований / С. Н. Кириллов, А. С. Шелудяков // Электросвязь. – 1999. – № 1. – С. 15–16.
98. Кловский, Д.Д. Теория электрической связи: Учебник для вузов / А.Г. Зюко, Д.Д. Кловский, В.И. Коржик, М.В. Назаров// Под ред. Д.Д. Кловского. – М.: Радио и связь, 1999. – 432 с.
99. Ковалгин, Ю. А. Цифровое кодирование звуковых сигналов : Учеб. пособие для вузов. / Ю. А. Ковалгин, Э. И. Вологдин // – СПб.: КОРОНА – ПРИНТ, 2004. – 240 с.
100. Коротаев Г. А. Эффективный алгоритм кодирования речевого сигнала на скорости 4,8 кбит/с и ниже / Г. А. Коротаев // Зарубежная радиоэлектроника. – 1996. – № 3. – С. 52 – 68.
101. Коротаев, Г. А. Анализ и синтез речевого сигнала методом линейного предсказания / Г. А. Коротаев // Зарубежная радиоэлектроника. – 1990. – № 3. – С. 31–51.

102. Коротаев Г. А. Некоторые аспекты линейного предсказания при анализе и синтезе речевого сигнала / Г. А. Коротаев // Зарубежная радиоэлектроника. – 1991. – № 7. – С. 3-31.
103. Корнеев, В. В. Современные микропроцессоры / В. В. Корнеев, А. В. Киселев // Москва: Нолидж, 2000. – 320с.
104. Косарев, Е.Л. Методы обработки экспериментальных данных / Е.Л. Косарев // Москва: ФИЗМАТЛИТ. 2008. – 208 с.
105. Кравченко, С. Н. Рекуррентный метод построения уравнений авторегрессии для векторных процессов / С. Н. Кравченко, А. И. Назаренко // Техническая кибернетика. – 1992. – № 1. – С. 122–129.
106. Куприянов, М. С. Цифровая обработка сигналов : процессоры, алгоритмы, средства проектирования / М. С. Куприянов, Б. Д. Матюшкин // СПб.: Политехника, 1998. – 592 с.
107. Лабутин, В. К. Модели механизмов слуха / В. К. Лабутин, А. П. Молчанов // Москва: Энергия, 1973. – 200 с.
108. Лагутенко, О. Современные модемы / О. Лагутенко // Москва: Эко-Трендз, 2005. – 346 с.
109. Ланнэ, А. А. Передача информации о состоянии фильтра-предсказателя с помощью спектральных пар / А. А. Ланнэ, Д. А. Улахович // Радиоэлектроника и связь. 1991. – № 1. – С. 37 – 43.
110. Ланнэ, А. А. Новая теория линейных спектральных корней // Цифровая обработка сигналов и ее применение: труды 3-й Международной конференции: труды / А. А. Ланнэ // Москва, 2000. – С. 118 – 125.
111. Лившиц, М. З. Широкополосный CELP – кодер с мультиполосным возбуждением и многоуровневым векторным квантованием по кодовой книге с реконфигурируемой структурой / М. З. Лившиц, М. Парфенюк, А. А. Петровский // Цифровая обработка сигналов № 2, 2005. – С. 20–35.
112. Липкин, И. А. Статистическая радиотехника. Теория информации и кодирования / И. А. Липкин // Москва: «Вузовская книга», 2002. – 216с.

113. Лобашев, А. И. Математические модели и алгоритмы функционирования систем кодирования речевых сообщений со структурно-параметрической адаптацией / А. И. Лобашев, А. А. Устинов, В. А. Тарусов // Труды учебных заведений связи. – СПб. : СПбГУТ, 2002. – № 168. – С. 12–17.
114. Макхоул, Д. Векторное квантование при кодировании речи / Д. Макхоул, С. Рукос, Г. Гиш // ТИИЭР. – 1985. – Т.73. – №11. – С. 19–61.
115. Малахов, А.Н. Кумулянтный анализ негауссовых случайных процессов и их преобразований / А.Н. Малахов // Москва: Сов.Радио, 1978. – 372 с.
116. Маркел, Дж. Д. Линейное предсказание речи / Д. Маркел, А. Х. Грей // Пер. с англ. под ред. Ю. Н. Прохорова и В. С. Звездина. – Москва: Связь, 1980. – 308 с.
117. Марпл-мл., С. Л. Цифровой спектральный анализ и его приложения / С. Л. Марпл-мл. // Москва: Мир, 1990. – 584 с.
118. Методы передачи изображений. Сокращение избыточности. Под редакцией Прэтта У. К. – Москва: Радио и связь, 1983. – 264 с.
119. Методы автоматического распознавания речи: в 2 кн. / под ред. У. Ли. – Москва: Мир, 1983.
120. Михайлов, В. Г. Измерение параметров речи / В. Г. Михайлов, Л. В. Златоустова // Москва: Радио и связь, 1987. – 168 с.
121. Мирошник, И. В., Никифоров В. О., Фрадков А. Л. Нелинейное и адаптивное управление сложными динамическими системами / И. В. Мирошник, В. О. Никифоров, А. Л. Фрадков // СПб.: Наука, 2000. – 549 с.
122. Назаров, М. В. Методы цифровой обработки и передачи речевых сигналов / М. В. Назаров, Ю. Н. Прохоров // Москва: Радио и связь, 1985. – 176 с.
123. Нейман, В. И. Интернет-телефония и перспективы ее развития / В. И. Нейман, Д. В. Селезнев // Электросвязь. – 2008. – № 1. – С. 6–9.
124. Никиас, Х.Л. Биспектральное оценивание применительно к цифровой обработке сигналов / Х.Л. Никиас // ТИИЭР. 1987. Т.75. - №7. – С. 5-30.
125. Никифоров, В. О. Адаптивное и робастное управление с компенсацией возмущений / В. О. Никифоров // СПб.: Наука, 2003. – 282 с.

126. Олифер В. Г., Олифер Н. А. Компьютерные сети. Принципы, технологии, протоколы: учебник для вузов. – 2-е изд./ В.Г. Олифер, Н.А. Олифер // СПб: Питер, 2005. – 864 с.

127. Павловец А. Н., Зубрыцки П., Петровский А. А. Гармоническая модель речевого сигнала : определение параметров и их квантование / А. Н. Павловец, П. Зубрыцки, А. А. Петровский // Доклады БГУИР.–2007–№ 4(20). – С. 19-34.

128. Патент № 2343564 РФ, МПК G10L 19/06. Способ адаптивного кодирования речевых сигналов на основе системы с переменной структурой / А. А. Афанасьев, Г.В. Богачев и др.; патентообладатель гос. казенное образовательное учреждение высш. проф. образования Академия ФСО России. – № 2006143249/09; заявл. 06.12.2006, опубл. 10.01.2009. Бюл. № 1.

129. Патент на полезную модель № 78977 РФ, МПК G10L 11/00. Анализатор основного тона и сигнала тон-шум / А.А. Афанасьев, О.О. Басов и др.; патентообладатель гос. казенное образовательное учреждение высш. проф. образования Академия ФСО России. – № 2008126595/22; заявл. 30.06.2008, опубл. 10.12.2008. Бюл. № 34.

130. Патент №2374653 РФ, МПК G01R23/16.Способ цифрового спектрального анализа сигналов / А.А. Афанасьев, М.Ю. Конышев и др.; патентообладатель гос. казенное образовательное учреждение высш. проф. образования Академия ФСО России.–№2008127698/28;заявл.07.07.2008, опубл. 27.11.2009. Бюл.№ 33.

131. Патент № 2400832 РФ, МПК G10L 19/08. Способ формирования сигнала возбуждения в низкоскоростных вокодерах с линейным предсказанием / А.А. Афанасьев, М.Ю. Конышев и др.; патентообладатель гос. казенное образовательное учреждение высш. проф. образования Академия ФСО России. – № 2008146294/09; заявл. 24.11.2008, опубл. 27.05.2010. Бюл. № 15.

132. Патент № 2394284 РФ, МПК G10L 19/08. Способ сжатия и восстановления речевых сигналов для систем кодирования с переменной скоростью передачи / А.А. Афанасьев, С.Р. Габдулгазиев и др.; патентообладатель гос. казенное образовательное учреждение высш. проф. образования Академия ФСО России. – № 2009110842/09; заявл. 24.03.2009, опубл.10.07.2010. Бюл. № 19.

133. Патент №2400831 РФ, МПК G10L 19/00. Способ выделения сегментов квазистационарности при анализе речевого сигнала в вокодерах с линейным предсказанием / А.А. Афанасьев, С.Р. Габдулгазиев и др.; патентообладатель гос. казенное образовательное учреждение высш. проф. образования Академия ФСО России. – №2009121174/09; заявл. 03.06.2009, опубл. 27.09.2010. Бюл.№27.

134. Патент № 2432624 РФ, МПК G10L 19/00. Способ уменьшения объема данных при широкополосном кодировании речевого сигнала / А.А. Афанасьев, В.Г.Трубицын и др.; патентообладатель гос. казенное образовательное учреждение высш. проф. образования Академия ФСО России. – №2010115915/09, заявл. 21.04.2010, опубл. 27.10.2011. Бюл. №30.

135. Патент № 2435214 РФ, МПК G06F 17/30, H03M 7/00. Способ быстрого поиска в кодовой книге при векторном квантовании / Афанасьев А.А., Габдулгазиев С.Р.; патентообладатель гос. казенное образовательное учреждение высш. проф. образования Академия ФСО России. – № 2010103261/08, заявл. 01.02.2010, опубл. 27.11.2011. Бюл. №33.

136. Патент № 2445719 РФ, МПК G10L 19/08. Способ улучшения восприятия синтезированной речи при реализации процедуры анализа через синтез в вокодерах с линейным предсказанием. / А.А. Афанасьев, М.В. Илюшини др.; патентообладатель гос. казенное образовательное учреждение высш. проф. образования Академия ФСО России. – №2010115914/08, заявл. 21.04.2010, опубл. 20.03.2012. Бюл. №8.

137. Патент № 2445718 РФ, МПК G10L 19/00. Способ выделения сегментов обработки речи на основе анализа корреляционных зависимостей в речевом сигнале/ А.А. Афанасьев, В.Г. Трубицын и др.; патентообладатель гос. казенное образовательное учреждение высш. проф. образования Академия ФСО России. – № 2010136618/08, заявл. 31.08.2010, опубл. от 20.03.2012. Бюл. №8.

138. Патент № 2459373, МПК H04L 29/06, H04L 12/56. Способ определения длины кадра передачи кодеков речевых сигналов на основе линейного предсказания в сетях с пакетной коммутацией на основе IP-протокола / А.А. Афанасьев, О.Н. Титов; патентообладатель гос. казенное образовательное учреждение высш. проф.

образования Академия ФСО России. – № 2010151566/08, заявл. 15.12.2010, опубл. 20.08.2012. Бюл. №23.

139. Патент № 2495504, МПК G10L 19/08. Способ снижения скорости передачи низкоскоростных вокодеров с линейным предсказанием / А.А. Афанасьев, А.П. Рыжков; патентообладатель гос. казенное образовательное учреждение высш. проф. образования Академия ФСО России. – № 2012126522/08, заявл. 25.06.2012, опубл. 10.10.2013. Бюл. №28.

140. Патент № 2504027, МПК G10L 19/09, G10L 19/038 Способ создания кодовой книги и поиска в ней при векторном квантовании данных / А.А. Афанасьев, А.П. Рыжков и др.; патентообладатель гос. казенное образовательное учреждение высш. проф. образования Академия ФСО России. – №2012128241/08, заявл. 03.07.2012, опубл. 10.01.2014. Бюл. №1.

141. Патент №2580796, МПК G10L 19/00, G06F 17/14, H04B 15/00. Способ (варианты) фильтрации зашумленного речевого сигнала в условиях сложной помеховой обстановки / А.А. Афанасьев, О.Н. Титов и др.; патентообладатель гос. казенное образовательное учреждение высш. проф. образования Академия ФСО России. – № 2015107227/08, заявл. 02.03.2015, опубл. 10.04.2016. Бюл. №10.

142. Патент № 2606566, МПК G10L 15/08, G10L 15/00. Способ и устройство классификации сегментов зашумленной речи с использованием полиспектрального анализа / А.А. Афанасьев, О.Н.Титов и др.; патентообладатель гос. казенное образовательное учреждение высш. проф. образования Академия ФСО России. – № 2014154081, заявл. 29.12.2014, опубл. 10.01.2017. Бюл. №1.

143. Патент № 2631968, МПК G10L 19/08 Способ низкоскоростного кодирования и декодирования речевого сигнала / А.А. Афанасьев, Е.С. Коробовский; патентообладатель гос. казенное образовательное учреждение высш. проф. образования Академия ФСО России. – № 2015127496, заявл. 08.07.2015, опубл. 29.09.2017. Бюл. №28.

144. Патрик, Э. Основы теории распознавания образов : Пер. с англ / Э. Патрик. – Москва: Сов. радио, 1980. – 408 с.

145. Петровский, А. А. Конспект лекций по дисциплине «Речевой интерфейс» / А. А. Петровский // Минск, 2007. – 292 с.
146. Петровский, А. А. Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением / А. А. Петровский, А. Борович, М. Парфенюк // Речевые технологии № 3, 2008. – С. 3–15.
147. Поляков, А. Н. Об одном из способов решения задачи определения оптимальных управляющих параметров системы низкоскоростной компрессии речевой информации / А. Н. Поляков // Телекоммуникации. 2008. №3.–С.15–18.
148. Попов, О. Б. Цифровая обработка сигналов в трактах звукового вещания / О. Б. Попов, С. Г. Рихтер // Учебн. пособие для вузов. – Москва: Горячая линия – Телеком, 2007. – 341 с.
149. Потапова, Р. К. Речь: коммуникация, информация, кибернетика: учеб. пособие для вузов / Р. К. Потапова // М. : Радио и связь, 1997. – 528 с.
150. Применение цифровой обработки сигналов. Под редакцией А. Оппенгейма. – М. : Мир, 1980. – 552 с.
151. Прокис, Дж. Цифровая связь / Дж. Прокис // Пер. с англ. под ред. Д. Д. Кловского. Москва: Радио и связь, 2000. – 800 с.
152. Прохоров, Ю. Н. Статистические модели и рекуррентное предсказание речевых сигналов/ Ю. Н. Прохоров // Москва: Радио и связь, 1984.– 240 с.
153. Процессоры цифровой обработки сигналов фирмы Texas Instruments. – М. : ЗАО SCAN, 1999, 75 с.
154. Рабинер, Л. Р. Цифровая обработка речевых сигналов / Л. Р. Рабинер, Р. В. Шафер // Москва: Радио и связь, 1981. – 496 с.
155. Радиовещание и электроакустика: учеб. для вузов / Под ред. Ю. А. Ковалгина. – М.: Радио и связь, 1998. – 792 с.
156. Радзишевский, А. Ю. Основы аналогового и цифрового звука / А. Ю. Радзишевский // Москва: Издательский дом «Вильямс», 2006. – 288 с.
157. Рамишвили, Г. С. Автоматическое опознавание говорящего по голосу / Г. С. Рамишвили // Москва: Радио и связь, 1981. – 224 с.

158. Рихтер, С. Г. Кодирование и передача речи в цифровых системах подвижной радиосвязи / С. Г. Рихтер // Учебное пособие для вузов. – Москва: Горячая линия – Телеком, 2010. – 304 с.
159. Сапожков, М. А. Речевой сигнал в кибернетике и связи / М. А. Сапожков // Москва : Государственное издательство литературы по вопросам связи и радио, 1963. – 450 с.
160. Сапожков, М. А. Электроакустика / М. А. Сапожков // Москва: Связь, 1978. – 272 с.
161. Сапожков, М. А. Вокодерная связь / М. А. Сапожков, В. Г. Михайлов // Москва: Радио и связь, 1983. – 248 с.
162. Свириденко, В. А. Анализ систем со сжатием данных / В. А. Свириденко // Москва: Связь, 1977. – 184 с.
163. Секунов, Н. Ю. Обработка звука на РС / Н. Ю. Секунов // СПб. : БХВ – Петербург, 2001.
164. Семенов, В. Ю. Новый метод вычисления линейных спектральных частот речевых сигналов, основанный на универсальном алгоритме решения трансцендентных уравнений / В. Ю. Семенов // Акустический вестник. 2002. Т. 5, № 4. С. 38 – 50.
165. Семенов, В. Ю. Новый подход к вычислению линейных спектральных частот речевых сигналов, основанный на свойстве межкадровой упорядоченности / В. Ю. Семенов // Акустический вестник. 2004. Т. 7, № 3. С. 55 – 64.
166. Сергиенко, А. Б. Цифровая обработка сигналов. Учебное пособие для вузов / А. Б. Сергиенко // С.-Пб.: Питер, 2002. – 608 с.
167. Ситняковский, И. В. Цифровая сельская связь / И. В. Ситняковский, В.И. Мейкшан // Под ред. М.Д. Венедиктова. – Москва: Радио и связь, 1994. – 248 с.
168. Скляр, Б. Цифровая связь. Теоретические основы и практическое применение, 2-е издание / Б. Скляр // Пер. с англ. – Москва: Издательский дом «Вильямс», 2003. – 1104 с.
169. Соболев, В. Н. Информационные технологии в синтетической телефонии / В. Н. Соболев // Монография – М. : ИРИАС, 2007. – 360 с.

170. Советов, Б. Я. Моделирование систем : учеб. для вузов по спец. «Автоматизир. системы обработки информ. и упр.». – 2-е изд., перераб. и доп / Б. Я. Советов // Москва: Высшая школа, 1998. – 319 с.
171. Современные речевые технологии. Сборник трудов IX сессии Российского акустического общества. – М. : ГЕОС, 1999. – 166 с.
172. Сперанский, В. С. Сигнальные микропроцессоры и их применение в системах телекоммуникаций и электроники : учебное пособие для вузов / В. С. Сперанский// Москва: Горячая линия-Телеком, 2008. – 168 с.
173. Справочник по теории автоматического управления. Под редакцией А.А. Красовского.– Москва: Наука, 1987. – 712с.
174. Тоцкий, А. В. Восстановление сигналов по оценкам биспектров в присутствии гауссовых и негауссовых помех / А. В. Тоцкий // Зарубежная радиоэлектроника. – 2002. – № 11. – С. 44–58.
175. Уидроу, Б. Адаптивная обработка сигналов / Б. Уидроу, С. Стирнз // Москва: Радио и связь, 1989. – 440 с. : ил.
176. Устинов, А. А. Стохастическое кодирование видео- и речевой информации: Моногр. Ч. 1 / А. А. Устинов // Под ред. проф. В. Ф. Комаровича. – СПб. : ВАС, 2005. – 220 с.
177. Фланаган, Д. Анализ, синтез и восприятие речи / Д. Фланаган // – Москва: Связь, 1968. – 396 с.
178. Хайкин, С. Нейронные сети: полный курс, 2-е издание. : Пер. с англ. / С. Хайкин // Москва: Издательский дом «Вильямс», 2006. – 1104 с.
179. Харатишвили, Н. Г. Цифровое кодирование с предсказанием непрерывных сигналов / Н. Г. Харатишвили // Москва: Радио и связь, 1986. – 140 с.
180. Харатишвили, Н. Г. Спутниковое телевидение. Новые методы передачи / Н. Г. Харатишвили // Москва: Радио и связь, 1993. – 244 с.
181. Цвикер, Э. Ухо как приемник информации /Э. Цвикер, Р. Фельдкеллер // Перевод с немецкого под редакцией Б. Г. Белкина.– М.: Связь,1971– 255с.
182. Цыпкин, Я. З. Информационная теория идентификации / Я. З. Цыпкин // М.: Наука. Физматлит, 1995. – 336 с.

183. Чистович, Л. А. Физиология речи. Восприятие речи человеком / Л. А. Чистович, А. В. Венцов, М. П. Гранстрем // Л., «Наука», 1976. – 388 с.
184. Чучупал, В. Я. Цифровая фильтрация зашумленных речевых сигналов / В. Я. Чучупал // М.: Вычислительный центр РАН, 1998. – 52 с.
185. Шалимов, И. А. Практикум по цифровой телефонии : учеб. пособие / И. А. Шалимов // Академия ФСБ России, 2008. – 344 с.
186. Шалимов, И. А. Теоретико-информационные принципы компрессии речевого сигнала на основе его квазипериодических свойств: диссертация доктора технических наук: 05.13.17. – Москва. – 2005. – 241 с. РГБ ОД, 71:06-5/528.
187. Шафер, Р. Цифровое представление речевых сигналов / Р. Шафер, Л. Рабинер // ТИИЭР, 1977. – Т.65, № 12.
188. Шелухин, О. И. Цифровая обработка и передача речи / О. И. Шелухин, Н.Ф. Лукьянцев // Москва: Радио и связь, 2000. – 456 с.
189. Atal, B. S. Predictive coding of speech at low bit rates/ B. S. Atal // IEEE Trans. On Comm. April 1982, 600 – 614.
190. Abboud, K. Wideband CELP Speech Coding / K. Abboud // Departament of Electrical Engineering, 1992. – 89 p.
191. Akaike, H. Likelihood of a model and information criteria / H. Akaike // Journal of Econometrics, 1981. – Vol. 16. – Pp. 3-14.
192. Afanasjev, A.A. A Working out of Speech Compression Adaptive Algorithms Based on the Systems with Variable Structure/ A.A. Afanasjev // Proceedings of International Workshop "Speech and computer" SPECOM'2001. – Moscow: MSLU, 2001. – P.136 – 138.
193. Chamberlain, M. A 600 bps MELP vocoder for use on HF channels / M. Chamberlain // MILCOM-2001. – 2001. – vol. 1. – P. 447-453.
194. Chang, W. W. A masking-threshold-adapted weighting filter for excitation search / W. W. Chang, C. T. Wang // IEEE Trans. on Speech and Audio process. – 1996. – Vol. 4. – № 2. – P. 124-131.

195. Chu Wai, C. Speech coding algorithms: Foundation and evolution of standardized coders / Chu Wai C. // John Wiley & Sons, Inc.; Hoboken, New Jersey: USA, 2003. – 558 p.
196. Cuperman, V. Robust vector quantization for noisy channels using soft decision and sequential decoding / V. Cuperman, F. H. Liu, P. Ho // ETT. – 1994. – vol. 5. – № 5. – P. 7-18.
197. Dall'Agnol, S. Performance of LSF vector quantizers for VSELP coders in noisy channels / S. Dall'Agnol, A. Alcaim, J. R. B. de Marca // ETT. – 1994. – Vol. 5. – № 5. – P.19-29.
198. Gersho, A. Concepts and paradigms in speech coding / A. Gersho // Speech Recognition and Coding. New Advances and Trends. NATO ASI Series. – Germany, 1995. – P. 370–390.
199. Goldberg, R. Practical Handbook of Speech Coders / R. Goldberg // L. Riek. CRC Press, Boca Raton, FL, 2000.
200. Griffin D. W. Multi-band excitation vocoder / D. W. Griffin // Cambridge: Massachusetts Institute of Technology, 1987. – 118 p.
201. Hasegawa-Johnson, M. Speech coding: fundamentals and applications / M. Hasegawa-Johnson, A. Alwan // Wiley Encyclopedia of Telecommunications, Edited by John G. Proakis, ISBN 0-471-36972, 2003. – 20 p.
202. Hayashi, S. Speech quality assessment of CS – ACELP/ S. Hayashi, A. Kataoka, T. Moriya, T. Kaneko // NTT Review. – 1996. – Vol. 8. – No. 4. – P. 36-41.
203. Huang, X. Spoken language processing: a guide to theory, algorithm and system development / X. Huang, A. Acero, H.-W. Hon // New Jersey: Prentice Hall, Inc., 2001. – 980 p.
204. Itakura, F. Line spectrum representation of linear predictive coefficients of speech signals/ F. Itakura // J. Acoust. Soc. America. – 1975. – vol. 57. – P.535.
205. ITU-T Recommendation P.800. Method for the subjective assessment of transmission quality, 1996.
206. ITU-T Recommendation G.729. Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic Code Excited Linear Prediction (CS-ACELP), 1996.

207. Jayant, N. Signal compression based on models of human perception / N. Jayant, J. Johnston, R. Safranek // *Proceedings of the IEEE*. – 1993. – vol. 81. – № 10. – P. 1385-1421.
208. Jean, F. R. Transparent quantization of speech LSP parameters based on KLT and 2-D-prediction / F. R. Jean, H. C. Wang // *IEEE Trans. on Speech and Audio process.* – 1996. – vol. 4. – № 1. – P. 60-66.
209. Juang, B. H. The past, present and future of speech processing / B. H. Juang // *IEEE Signal Processing Magazine*. – 1998. – Vol. 15. – No.3. – P. 24–48.
210. Kataoka, A. LSP and gain quantization for CS – ACELP speech coder / A. Kataoka, S. Hayashi, T. Moriya, J. Ikeda // *NTT Review*. – 1996. – Vol.8. – № 4. – P. 30-35.
211. Kataoka, A. Basic algorithm of conjugate-structure algebraic CELP (CS – ASELP) speech coder / A. Kataoka, S. Hayashi, T. Moriya, S. Kurihara, K. Mano // *NTT Review*. – 1996. – Vol. 8. – No. 4. – P. 24–29.
212. Kitawaki, N. An 8-kbit/s speech coding method (CS – ACELP) standardized by ITU / N. Kitawaki // *NTT Review*. – 1996. – Vol. 8. – No. 4. – P. 16–23.
213. Kleijn, W.B. Encoding speech using prototype waveforms /W.B. Kleijn// *IEEE Trans. on Speech and Audio Processing*.–1993.–Vol.1.–№ 4.– P. 386-399.
214. Koestoer, N. P. Robust linear prediction analysis for low bit-rate speech coding /N. P. Koestoer//DPh dissertation.–Brisbane:Griffith University, 2002.–132 p.
215. Luttrell, S. P. Hierarchical vector quantization / S. P. Luttrell // *IEE Proceedings (London)*, 1989, vol. 136 (part 1), P. 405-413.
216. Martin R., Malah D., Cox R. V., Accardi A. J. A noise reduction preprocessor for mobile voice communication / R. Martin, D. Malah, R. V. Cox, A. J. Accardi // *EURASIP J. of Applied Signal Processing*. – 2004. – № 8. – P. 1046-1058.
217. McCree, A. V. A mixed excitation LPC vocoder model for low bit rate speech coding / A. V. McCree, T. P. Barnwell // *IEEE Trans. Speech and Audio Processing*. – 1995. – vol. 3. – P. 242-250.

218. Moreau, N. Dymarski P. Selection of excitation vectors for the CELP coders/ N. Moreau , P. Dymarski // IEEE Transactions on speech and audio processing. – 1994. – Vol. 2. – № 1. – Part 1. – P. 29-40.
219. Noll, P. Speech And Audio Coding For Multimedia Communications / P. Noll // Proc. of the IEEE. Vol. 81, No. 10, PP. 1385–1422.
220. Painter, T. Perceptual coding of digital audio / T. Painter, A. Spanias // IEEE. – 2000. – Vol. 88. – № 4. – P. 451–513.
221. Palival, K. K.. Efficient vector quantization of LPC parameters at 24 bits/frame / K. K. Palival, B. S. Atal // IEEE Trans. on Acoustics, Speech and Signal Processing. – 1993. – Vol. 1. – № 1. – P. 3-14.
222. Pan, J. Vector quantization of speech line spectrum pair parameters and reflection coefficients / J. Pan, T. R. Fischer // IEEE Trans. on Speech and Audio process. – 1998. – vol. 6. – № 2. – P.106-115.
223. Ribeiro, C. M. Speaker adaptation in a phonetic vocoding environment / C. M. Ribeiro, I.M. Trancoso // IEEE Workshop on Speech Coding Proceedings. – 1999. – P. 64-66.
224. Rizvi, A. Predictive vector quantizer using constrained optimization /A. Rizvi, N. M. Nasrabadi // IEEE Signal processing letters. – 1994. – Vol. 1. – № 1. – P. 15-18.
225. Sinha, D. Unequal error protection methods for perceptual audio coders / D. Sinha // Proc. ICASSP, 1999. – Vol. 5. – P. 2423–2426.
226. Spanias, A. Speech Coding : A Tutorial Review / A. Spanias // Proceedings of the IEEE. Vol. 82, No. 10, 1994.
227. Speech Coding. New Advances and Trends. NATO ASI Series. – Germany, 1998. – 445 p.
228. Speech enhancement: signals and communication technology / Editors Benesty J., Makino S., Chen J. // Springer, 2005. – 406 p.
229. Speech Recognition and Coding. New Advances and Trends. NATO ASI Series. – Germany, 1995. – 464 p.

230. Supplee, L. M. MELP: The new Federal Standard at 2400 bps / L. M. Supplee, R.P. Cohn, J. S. Collura, A. V. McCree // IEEE ICASSP-97 Conference, Munich Germany. – P.1591-1594.
231. Thomson, D. L. Selective modeling of the LPC residual during unvoiced frames; white noise or pulse excitation / D. L. Thomson, D. P. Prezias // IEEE Int. Conf. on Acoustics, Speech, and Signal Processing. – Tokyo, 1986. – P. 3087-3090.
232. Trancoso, I. M. An overview of different trends on CELP coding / I. M. Trancoso // Speech Recognition and Coding. New Advances and Trends. NATO ASI Series. – Germany, 1995. – P. 351–367.
233. Vary, P., Martin R. Digital speech transmission: enhancement, coding and error concealment / P. Vary, R.Martin // John Wiley & Sons, Inc., 2006. – 376 p.
234. Voice activity detection. ETSI/GSM Recommendation 06.32.
235. Voran, S. Objective Estimation of Perceived Speech Quality /S. Voran// IEEE Transactions on Speech and Audio Processing, 1999. – P. 286–310.
236. Wong, W. T. K. Low rate speech coding for telecommunications / W. T. K. Wong, R. M. Mack, B. M. G. Cheethman, X. Q. Sun // BT Technol J. – 1996. – Vol. 14. – No. 1. – P. 28–43.
237. Yang, W. Enhanced Modified Bark Spectral Distortion (EMBSD) : An Objective Speech Quality Measure Based On Audible Distortion And Cognition Model / A Dissertation of the Requirement for the Degree Doctor of Philosophy – May, 1999.
238. Компания "Texas Instruments" [Электронный ресурс]: [сайт]. – Режим доступа: <http://www.ti.com>. – Дата обращения: 5.06.2018.
239. Научно-производственный центр "ЭЛВИС" [Электронный ресурс]: [сайт]. – Режим доступа: <http://www.elvees.ru>. – Дата обращения: 5.06.2018.
240. Научно-технический центр "Модуль" [Электронный ресурс]: [сайт]. – Режим доступа: <http://www.module.ru>. – Дата обращения: 5.06.2018.
241. Компания "MathWorks" [Электронный ресурс]: [сайт]. – Режим доступа: <http://www.mathworks.com>. – Дата обращения: 5.06.2018.

ПРИЛОЖЕНИЕ 1 Особенности формирования тестовых фраз для проверки разработанных методов и алгоритмов

Тестовые фразы получены при произношении звуков мужской слитной речи (3 диктора) с использованием микропроцессорной техники и программного обеспечения. АЦП РС реализовано на частоте дискретизации 8 и 22,05 кГц с использованием 8 и 16-битного квантователя. Перечень тестовых фраз определен в соответствии с ГОСТ Р 51061-97 "Системы низкоскоростной передачи речи по цифровым каналам. Параметры качества речи и методы измерений". Подчеркнуты соответствующие вокализованные и невокализованные участки РС, используемые для дальнейшего анализа.

1. "Если хочешь быть здоров, советует Татьяна Илье, чисть зубы пастой "Жемчуг "!
2. "В клумбах сочинской здравницы «Пуща», сообщает нам автоинспектор, обожгли шихту".
3. "Тропический какаду — это крупный попугай? Ты не злословишь?"
4. "Так ты считаешь, что техникой мы обеспечены на весь сезон?"
5. "Раз. Эти жирные сазаны ушли под палубу".

Особенности формирования тестовых зашумленных записей

Длительность записей устанавливалась 10 секунд при обработке на сегменте локальной стационарности, не превышающем 40 мс при различных частотах дискретизации 8000, 11025, 16000, 22050, 32000, 44100, 64000 кГц.

Используемые тестовые фразы согласно [84] содержат информацию обо всех фонемах русского языка, включая информацию и об аллофонах.

Для проверки чувствительности модели данные тестовые фразы, аддитивно зашумлялись различными видами акустических воздействий (Таблица 4.1) квазистационарной природы (КБГШ, двигатель, город, ветер и др.) при различных значениях ОСШ. Запись тестовых фраз и акустических воздействий производилась с

частотой дискретизации 8 кГц с использованием 16-битного квантователя при нормировании входного РС, учитывающей количество бит на отсчет (76800 отсчетов), обработка РС ведется на сегментах длительностью 20 и 32 мс (160 и 256 отсчетов соответственно). Моделирование выполнялось в среде программирования Math-Works MATLAB R2012a.

Формирование тестового материала, необходимого для оценки функционирования, осуществляется согласно следующим пунктам:

1) Использование тестовых фраз согласно [84]. Данные тестовые фразы полностью характеризуют вариабельность русского языка, при этом учитывают его параметрические и статистические свойства.

2) Запись тестовых фраз осуществляется от 13 дикторов, 8 из которых мужского пола: 2 до 20 лет, 2 от 20 до 35 лет, 2 старше 35 лет, и 5 женского пола: 3 до 35 лет, 2 старше 35 лет.

3) Запись осуществляется в условия отсутствия зашумления.

4) Запись акустических воздействий производится от реальных источников длительностями по 20 секунд при различных частотах дискретизации (для случаев квазибелого гаусовского шума, шумовой смеси и тонового шума – формирование в программной среде MATLAB).

5) Нормировка тестовых фраз для всех дикторов и всех записей (средняя мощность всех чистых сигналов равна).

6) Нормировка записей шумовых воздействий (средняя мощность всех шумовых записей равна средней мощности чистых сигналов).

7) Зашумление тестовых фраз осуществляется аддитивно в программной среде MATLAB, при этом случайным образом выбирается отрезок длительностью 10 секунд из записи акустического сигнала с добавлением мультипликативного коэффициента уровня с шагом 1 дБ.

ПРИЛОЖЕНИЕ 2 Акты внедрения

УТВЕРЖДАЮ

Командир войсковой части

В.И. Лащук

« 2 » марта 2013 г.

Акт

о внедрении результатов диссертационной работы
Афанасьева Андрея Алексеевича,
выполненной на специальную тему

Комиссия в составе: председателя
Щербакова С.А., и членов

Кондрашина А.А.,
Силакова Н.П.,
Григораш М.В. составила настоящий акт о том, что диссертационная работа
Афанасьева А.А. на специальную тему рассматривает
вопросы по прикладным направлениям, которые связаны с
совершенствованием устройств обработки речевых сигналов и исследуются с
целью обеспечения функционального развития соответствующих устройств
систем связи.

Предложенные в работе Афанасьева А.А. научно-технические
решения:

- методика обработки речевого сигнала на основе учета взаимосвязи параметров его декомпозиции при линейном предсказании;
- методика снижения скорости передачи в низкоскоростных системах кодирования речи;
- методика уменьшения избыточности речевого сигнала при его обработке в системах связи;
- предложения по реализации данных методик в системах связи специального назначения;

имеют теоретическую ценность и практическую значимость и их целесообразно использовать при совершенствовании устройств обработки речевых сигналов в системах автоматического засекречивания информации.

Председатель комиссии:

Члены комиссии:

С.А. Щербаков

А.А. Кондрашин

Н.П. Силаков

М.В. Григораш

УТВЕРЖДАЮ

Генеральный директор
ООО НТЦ «Фобос-НТ»
А.В. Гусаров



31 мая 2018 г.

Акт

об использовании результатов диссертационной работы

Афанасьева Андрея Алексеевича

"Модели и методы анализа и обработки речевого сигнала в системах связи"
на соискание ученой степени доктора технических наук

«31» мая 2018 г.

Комиссия в составе: председателя комиссии заместителя генерального директора - директора департамента Орлика Сергея Владимировича и члена комиссии начальника отдела Крючкова Сергея Васильевича составила настоящий акт о том, что результаты диссертационной работы Афанасьева А.А.:

- модель системы фильтрации речевого сигнала, отличающаяся применением полиспектрального анализа и позволяющая разработать методы шумоподавления в речевом сигнале в широком диапазоне значений отношения сигнал-шум;

- технологии адаптивной цифровой фильтрации речевого сигнала в условиях акустического зашумления, отличающиеся применением полиспектрального анализа и позволяющие осуществлять качественное шумоподавление при низких отношениях сигнал-шум;

- метод выделения сегментов анализа речевого сигнала различной длительности на однородных участках речевого сигнала, позволяющий осуществлять синтез речевого сигнала при фиксированных значениях параметров передаточной функции голосового тракта;

- алгоритмы анализа и обработки речевого сигнала, позволяющие повысить характеристики функционирования телекоммуникационных устройств в системах инфокоммуникаций в условиях наличия акустических шумов, отличающиеся от известных учетом особенностей формирования и обработки речевого сигнала, применением результатов полиспектрального анализа с целью шумоподавления, устраняющие противоречия между особенностями речеобразования и существующими в современных системах связи технологиями обработки речевого сигнала

имеют научную значимость и практическую ценность и используются при совершенствовании устройств обработки речевого сигнала в системах телекоммуникаций.

Председатель комиссии:

С.В.Орлик

Члены комиссии:

С.В.Крючков

УТВЕРЖДАЮ

Директор

ООО «Ставсервис – Связь»

 В.П. Шутров
 04 июня 2018 г.

Акт

об использовании результатов диссертационной работы

Афанасьева Андрея Алексеевича

"Модели и методы анализа и обработки речевого сигнала в системах связи"
 на соискание ученой степени доктора технических наук

« 4 » _____ июня 2018 г.

Комиссия в составе: председателя — заместителя директора Шмелёва Анатолия Ивановича и членов начальника центра защиты информации Бржезовского Кирилла Александровича составила настоящий акт о том, что результаты диссертационной работы Афанасьева А.А.:

- модель обработки речевого сигнала, основанная на учете взаимозависимости элементов декомпозиции речевого сигнала при его низкоскоростном кодировании на основе метода линейного предсказания, позволяющая осуществить снижение мощности пространства представления сигналов возбуждения;

- метод обработки речевого сигнала, основанный на учете взаимозависимости элементов декомпозиции РС при линейном предсказании, позволяющий уменьшать среднюю скорость передачи в канале связи при сохранении качественных показателей синтезированной речи;

- предлагаемые технические решения по непрерывной аутентификации диктора для доступа к ресурсу пропускной способности сети связи

имеют научную значимость и практическую ценность и используются при совершенствовании устройств телекоммуникаций и специальных алгоритмов разграничения доступа в системах инфокоммуникаций.

Председатель комиссии:

А.И. Шмелёв

Члены комиссии:

К.А. Бржезовский

УТВЕРЖДАЮ



2003г.

АКТ РЕАЛИЗАЦИИ

результатов диссертационной работы Афанасьева Андрея Алексеевича

Комиссия в составе: председателя комиссии –

кандидата технических наук, доцента,

Тихалева Ю.В., членов комиссии:

Ма-

карова В.Н.;

кандидата технических наук, до-

цента

Лисичкина В.Г. составила настоящий акт о том, что результа-

ты диссертационной работы

Афанасьева А.А., а именно:

- обоснование перехода к цифровой обработке сигналов как тенденции развития ИТКС,
- состояние цифровой обработки сигналов на этапе совершенствования ИТКС СНА,
- перспективы применения цифровой обработки сигналов в ИТКС СНА,

имеют научную и практическую ценность, внедрены в учебный процесс Академии и отражены в научно-исследовательской работе "Разработка научно-методического обеспечения учебных дисциплин "Цифровая обработка сигналов", "Микропроцессоры и цифровая обработка сигналов", шифр "Цифра 7", №6293 бр (акт приемки 14/3/37, акт внедрения 14/3/36 от 22.07.03).

Председатель комиссии:

Handwritten signature of Tikhalev Yu.V.

Тихалев Ю.В.

Члены комиссии:

Handwritten signature of Makarov V.N.

Макаров В.Н.

Handwritten signature of Lisichkin V.G.

Лисичкин В.Г.

УТВЕРЖДАЮ

Проректор по научной работе
ФГБОУ ВПО «РГРТУ»



Кириллов С.Н.

14 декабря 2012 г.

Акт

о внедрении результатов диссертационной работы
Афанасьева Андрея Алексеевича, выполненной на специальную тему
и представляемой на соискание ученой степени доктора технических наук

«14» декабря 2012 г.

г. Рязань

Комиссия в составе: председателя проректора по научной работе Федерального государственного бюджетного образовательного учреждения высшего профессионального образования "Рязанский государственный радиотехнический университет", д.т.н., профессора Кириллова Сергея Николаевича и членов: доцента кафедры Радиоуправления и связи Дмитриева Владимира Тимуровича, доцента кафедры Радиоуправления и связи Слесарева Алексея Сергеевича, Старшего преподавателя кафедры радиоуправления и связи Косткина Ивана Вячеславовича составила настоящий акт о внедрении результатов диссертационной работы Афанасьева Андрея Алексеевича на специальную тему.

Комиссия подтверждает, что предложенные автором в работе методики изменения длительности сегмента анализа речевого сигнала и нахождения взаимозависимостей элементов его декомпозиции при низкоскоростном кодировании, а также алгоритмы обработки речевых сигналов, разработанные на их основе, имеют научную значимость. Научные результаты, полученные в работе, реализованы и использованы в учебном процессе кафедры в рамках преподавания дисциплины «Кодеки сигналов в МТКС».

Председатель комиссии:
Проректор по научной работе
ФГБОУ ВПО «РГРТУ»
д.т.н., профессор

Кириллов С.Н.

Члены комиссии:
Доцент каф. РУС, к.т.н.

Дмитриев В.Т.

Доцент каф. РУС, к.т.н.

Слесарев А.С.

Старший преподаватель каф. РУС

Косткин И.В.

УТВЕРЖДАЮ

Начальник Академии



З.В. Мизеров
З.В. Мизеров

20 февраля 2013 г.

Акт

об использовании результатов диссертационной работы
на соискание ученой степени доктора технических наук
Афанасьева Андрея Алексеевича

Комиссия в составе: председателя –

кандидата технических наук

Куцакина А.И. и членов:

кандидата технических наук доцента

Тихалёва Ю.В.,

кандидата технических наук доцента

Богданова Н.Г. констатирует, что результаты исследования Афанасьева А.А. на специальную тему, а именно *методика адаптивного выделения сегментов анализа речевого сигнала различной длины, методика выявления взаимозависимостей элементов декомпозиции речевого сигнала в процессе его обработки на основе метода линейного предсказания, методика уменьшения избыточности речевого сигнала при его обработке в системах связи, методика очистки речевого сигнала от шумов в процессе его обработки* имеют научную значимость и практическую ценность и использованы в образовательном процессе Академии ФСО России на кафедре "Электроника и теория связи" при разработке и проведении занятий курса повышения квалификации сотрудников ФСО России "Цифровая обработка сигналов в современных информационно-телекоммуникационных системах".

Методика адаптивного выделения сегментов анализа речевого сигнала различной длины, предложенная Афанасьевым А.А., использована при разработке лекции № 1 "Задачи и методы кодирования речевого сигнала" по теме № 4 "Цифровая обработка речевых сигналов и изображений" курса повышения квалификации сотрудников ФСО России "Цифровая обработка сигналов в современных информационно-телекоммуникационных системах".

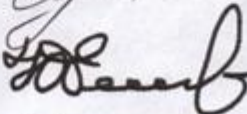
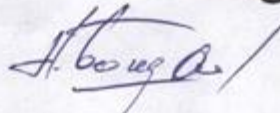
Методика выявления взаимозависимостей элементов декомпозиции речевого сигнала в процессе его обработки на основе метода линейного предсказания, предложенная Афанасьевым А.А., использована при разработке лекции № 3 "Линейное (гибридное) кодирование речевого сигнала" по теме № 4 "Цифровая обработка речевых сигналов и изображений" курса повышения квалификации сотрудников ФСО России "Цифровая обработка сигналов в современных информационно-телекоммуникационных системах".

Методика уменьшения избыточности речевого сигнала при его обработке в системах связи, предложенная Афанасьевым А.А., использована при разработке самостоятельной работы под руководством преподавателя "Непосредственное и вокодерное кодирование речевого сигнала" по теме № 4 "Цифровая обработка речевых сигналов и изображений" курса повышения квалификации сотрудников ФСО России "Цифровая обработка сигналов в современных информационно-телекоммуникационных системах".

Методика очистки речевого сигнала от шумов в процессе его обработки, предложенная Афанасьевым А.А., использована при разработке лекции № 1 "Задачи и методы кодирования речевого сигнала" по теме № 4 "Цифровая обработка речевых сигналов и изображений" курса повышения квалификации сотрудников ФСО России "Цифровая обработка сигналов в современных информационно-телекоммуникационных системах".

Председатель комиссии:

Члены комиссии:

 А.И. Куцакин
 Ю.В. Тихалёв
 Н.Г. Богданов