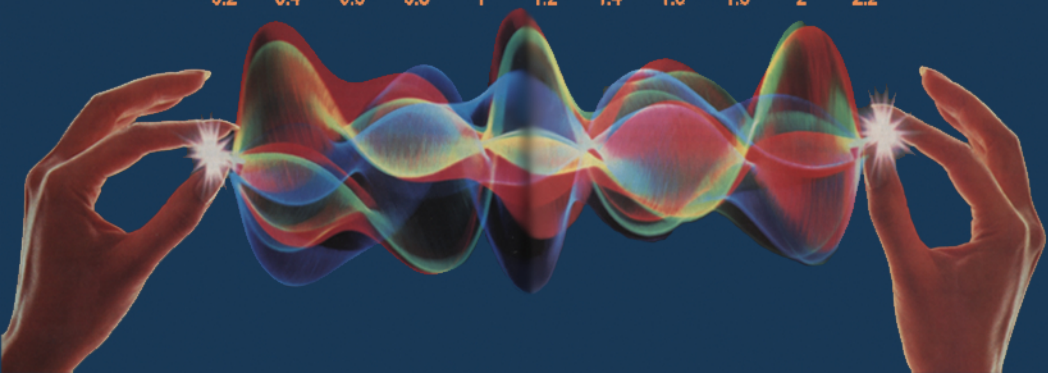
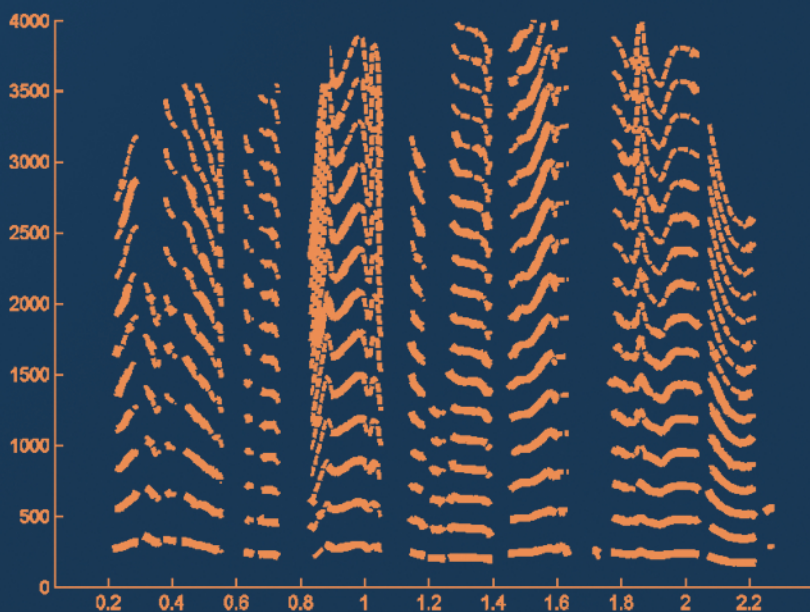


АНАЛИЗАТОРЫ РЕЧЕВЫХ И ЗВУКОВЫХ СИГНАЛОВ:

методы, алгоритмы и практика

(с MATLAB примерами)

под редакцией д.т.н. профессора Петровского А.А.



под редакцией д.т.н. профессора Петровского А.А.

**Анализаторы речевых и звуковых сигналов:
методы, алгоритмы и практика
(с MATLAB примерами)**

Минск
«Бестпринт»
2009

Анализаторы речевых и звуковых сигналов: методы, алгоритмы и практика (с MATLAB примерами) // под редакцией д.т.н. профессора Петровского А.А. – Минск: Бестпринт, 2009 – 456 с., – ISBN 978-985-6963-04-2.

В данной книге обобщен многолетний опыт сотрудников НИЛ 3.1 «Мультипроцессорные системы реального времени» БГУИР проектирования различных компонентов систем мультимедиа таких как субполосное кодирование аудио- и речевых сигналов, подавление шумов окружающей среды в речевом сигнале, компенсация акустического и электрического эхо-сигналов, и т.д. Главные направления исследований в данной области заключаются в получении новых функциональных свойств, повышении качества алгоритмов обработки, а также в увеличении эффективности систем мультимедиа. Параллельно с теоретическими результатами, показаны практические реализации на базе современных технологий таких, как реконфигурируемые и программируемые схемы, цифровые процессоры обработки сигналов.

Для научных и инженерно-технических работников, специализирующихся в области проектирования систем мультимедиа, аспирантов и студентов специальностей «Электронные вычислительные средства», «Вычислительные машины, системы и сети».

Рекомендовано Советом УО «Белорусский государственный университет информатики и радиоэлектроники», протокол №4 от 24 декабря 2009г.

Рецензенты:

д.т.н., профессор Голенков В.В.

д.т.н. Лобанов Б.М.

СОДЕРЖАНИЕ

Список авторов	8
Предисловие.....	9
1 Субполосный анализ и его эффективность.....	13
<i>Петровский Ал.А.</i>	
1.1 Введение в банки фильтров	13
1.2 Оценка эффективности субполосного кодирования.....	20
1.3 Список использованных источников	23
Часть 1 Методы субполосного анализа	
2 Введение в цифровые банки фильтров	25
<i>Вашкевич М.И., Парфенюк М., Петровский А.А.</i>	
2.1 Введение.....	25
2.2 Изменение частоты дискретизации	26
2.3 Построение банков фильтров на основе модуляции.....	29
2.4 Список использованных источников	48
3 Параунитарные банки фильтров на основе алгебры кватернионов	49
<i>Парфенюк М., Петровский А.А.</i>	
3.1 Решетчатые структуры параунитарных банков фильтров.....	50
3.2 Регулярность и эффекты квантования коэффициентов.....	51
3.3 Кватернионы и ортогональные матрицы	52
3.4 Решетчатые структуры ПУБФ на основе кватернионов.....	56
3.5 Примеры синтеза.....	63
3.6 Заключение.....	71
3.7 Список использованных источников	71
4 Анализ звуковых сигналов с адаптивным частотно-временным планом	74
<i>Петровский Ал.А.</i>	
4.1 Вейвлет преобразование.....	74
4.2 Методы декомпозиция ПДВП.....	84
4.3 Согласованная подгонка на основе адаптивного ПДВП.....	94
4.4 Список использованных источников	97
5 Дискретное преобразование Фурье с неравномерным частотным разрешением: определение, свойства, вычислительные аспекты.....	99
<i>Борович А., Парфенюк М., Петровский А.А.</i>	
5.1 Введение.....	99
5.2 Определение WDFT	100
5.3 Обратное преобразование WDFT.....	101
5.4 WDFT с избыточным синусоидальным базисом	104

5.5	Вычислительные аспекты WDFT	110
5.6	Аппроксимация шкалы барков	111
5.7	Комментарии о частотном разрешении WDFT	112
5.8	Список использованных источников	114
6	Мгновенный гармонический анализ звуковых и речевых сигналов	115
	<i>Азаров И.С., Петровский А.А.</i>	
6.1	Введение	115
6.2	Мгновенные гармонические параметры и классические методы их оценки	116
6.3	Оценка мгновенных гармонических параметров при помощи специального фильтра анализа	121
6.4	Сравнение точности оценки гармонических параметров различными методами	128
6.5	Список использованных источников	132
7	Психоакустические модели для систем мультимедиа на основе субполосной обработки сигналов	134
	<i>Петровский Ал.А., Парфенюк М.</i>	
7.1	Введение	135
7.2	ПДВП, согласованное со шкалой критических частотных полос	136
7.3	Процедура расчета порогов маскирования в вейвлет области	142
7.4	Перцептуальная обработка в домене 4-х полосного параунитарного банка фильтров на основе кватернионов	147
7.5	Психоакустическая модель на основе WDFT	153
7.6	Список использованных источников	161
Часть 2. Алгоритмы анализа и кодирования речевых и звуковых сигналов в системах мультимедиа		
8	Векторное квантование и компрессия звуковых сигналов	163
	<i>Лившиц М.З., Павловец А.Н., Лихачев Д.С., Петровский А.А.</i>	
8.1	Понятие векторного квантования	163
8.2	Меры искажений, используемые в векторном квантовании	167
8.3	Формирование кодовой книги	169
8.4	Векторное квантование с расщеплением	171
8.5	Многоуровневое векторное квантование	172
8.6	Векторное квантование коэффициентов линейного предсказания	172
8.7	Квантование амплитуд гармоник речевого сигнала с использованием закономерностей психоакустики	180
8.8	Список использованных источников	188
9	Применение гибридного параметрического описания для кодирования звуковых и речевых сигналов	190
	<i>Петровский Ал.А., Азаров И.С., Петровский А.А.</i>	
9.1	Введение	190
9.2	Общая структура системы гибридного описания	191
9.3	Синусоидальное и гармоническое представления сигнала	192
9.4	Оценка мгновенных гармонических параметров	194

9.5	Моделирование переходных составляющих сигнала методом согласованной подгонки	202
9.6	Применение гибридной системы анализа	205
9.7	Заключение	207
9.8	Список использованных источников	207
10	Компрессия речевого сигнала на основе синусоидальной модели с антропоморфической обработкой	211
	<i>Лихачев Д.С., Петровский А.А.</i>	
10.1	Задача антропоморфической обработки речевого сигнала применительно к системам компрессии на основе синусоидальной модели	211
10.2	Антропоморфическая обработка речевого сигнала на основе моделей слуха человека	214
10.3	Особенности построения системы компрессии речевого сигнала на основе синусоидальной модели с антропоморфической обработкой	225
10.4	Список использованных источников	230
11	Широкополосный CELP-кодер с мультиполосным возбуждением по банку кодовых книг с реконфигурируемой архитектурой	234
	<i>Лившиц М.З., Парфенюк М., Петровский А.А.</i>	
11.1	Модель широкополосного кодера	234
11.2	Психоакустическая модель для оценки перцептуальной значимости субполос	239
11.3	Математическая модель перцептуального CELP-кодера с реконфигурируемым банком кодовых книг	245
11.4	Препроцессор повышения качества речи	252
11.5	Математическая модель комбинированной системы компрессора широкополосных речевых сигналов с препроцессором редактирования шума	254
11.6	Список использованных источников	258
12	Гибридное кодирование речевых сигналов на основе ДПФ синхронизированного с изменением частоты основного тона	261
	<i>Павловец А.Н., Петровский А.А.</i>	
12.1	Введение	261
12.2	Гибридная модель речевого сигнала	262
12.3	Определение параметров модели	264
12.4	Синтез речи по параметрам модели "гармоники + шум"	275
12.5	Гибридный вокодер со скоростью передачи 2,4/4,8 кбит/с, основанный на декомпозиции речевого сигнала на периодическую и аperiodическую компоненты	276
12.6	Список использованных источников	282
13	Текстозависимая конверсия голоса на основе декомпозиции речевого сигнала на гармоники, шум и переходные фреймы	285
	<i>Павловец А.Н., Лившиц М.З., Лихачев Д.С., Петровский А.А.</i>	
13.1	Задача конверсии голоса	285
13.2	Декомпозиция речевого сигнала на гармоническую составляющую, шум и переходные фреймы	287
13.3	Реализация системы конверсии голоса	288

13.4	Экспериментальные результаты	296
13.5	Список использованных источников	298
14	Дискретное преобразование Фурье с неравномерным частотным разрешением в перцептуальных системах редактирования шума и компрессии звуковых сигналов	300
	<i>Борович А., Парфенюк М., Петровский Ал.А.</i>	
14.1	Дискретное преобразование Фурье с неравномерным частотным разрешением	301
14.2	Новая схема обработки сигналов в области коэффициентов WDFT	304
14.3	Схема подавления шума в речевом сигнале на основе спектрального взвешивания	305
14.4	Оценка порога маскирования и спектральной плотности мощности шума	309
14.5	Оценка качества системы подавления шума	316
14.6	Компрессия звуковых сигналов в области коэффициентов WDFT	321
14.7	Список использованных источников	329
15	Проектирование банка фильтров для слуховых аппаратов с использованием частотного растяжения и объединения субполос.....	331
	<i>Вашкевич М.И., Парфенюк М., Петровский А.А.</i>	
15.1	Введение.....	331
15.2	Косинусно-модулированный банк фильтров с деформированной частотной осью	333
15.3	Объединение субполос	342
15.4	Неравнополосный косинусно-модулированный банк фильтров с частотной деформацией и объединением субполос на основе фазовых звеньев без умножителей	345
15.5	Список использованных источников	350
16	Масштабируемый компрессор звуковых сигналов для IP технологий на основе перцептуально адаптируемого пакетного дискретного вейвлет преобразования	352
	<i>Петровский Ал.А.</i>	
16.1	Введение.....	353
16.2	Перцептуальный аудиокодер.....	353
16.3	Динамическая трансформация дерева ПДВП под заданный вычислительный ресурс.....	355
16.4	Масштабируемый компрессор звуковых сигналов для IP технологий	365
16.5	Список использованных источников	372
Часть 3 Аппаратная реализация анализаторов речевых и звуковых сигналов		
17	ПДВП - процессор на лестничных структурах.....	374
	<i>Родионов М.М., Петровский Ал.А.</i>	
17.1	Введение.....	374
17.2	Алгоритм ПДВП.....	375
17.3	Полифазная реализация банка фильтров	376
17.4	Факторизация полифазной матрицы. Переход к лестничной структуре банка.....	378
17.5	Двухканальный банк фильтров анализа и синтеза на лестничных структурах.....	381
17.6	Вопросы реализации банка фильтров на арифметике с фиксированной запятой	383
17.7	Поточная архитектура процессора ПДВП	389
17.8	Операционный блок процессора	392
17.9	Список использованных источников	396

18	Вычислительная платформа на основе ЦПОС TMS320C6713 и реконфигурируемой логики для мультимедиа приложений реального времени	397
	<i>Лившиц М.З., Петровский Ал. А., Качинский М.В., Станкевич А.В., Петровский А.А.</i>	
18.1	Вычислительная платформа для мультимедиа приложений реального времени.....	397
18.2	Система кодирования широкополосных речевых сигналов с препроцессором редактирования шума	403
18.3	Реализация кодера звука на базе ПДВП с динамической трансформацией частотно-временного плана на вычислительной платформе в реальном времени.....	420
18.4	Список использованных источников	426
19	Аппаратный ускоритель на базе ПЛИС с интерфейсом USB 2.0	430
	<i>Станкевич А.В., Петровский Ал.А., Ключ В.Б., Качинский М.В.</i>	
19.1	Выбор варианта реализации аппаратного ускорителя	430
19.2	Структура аппаратного ускорителя.....	431
19.3	Программное обеспечение	434
19.4	Исследование скорости обмена по интерфейсу USB 2.0	436
19.5	Заключение.....	437
19.6	Список использованных источников	437
20	ПДВП-анализатор звуковых сигналов на аппаратном ускорителе на базе ПЛИС.....	438
	<i>Петровский Ал.А., Станкевич А.В.</i>	
20.1	Введение.....	438
20.2	Аппроксимация критического частотно-временного плана	440
20.3	ПДВП-процессор реального времени	441
20.4	Параллельно-поточная архитектура процессора ПДВП.....	445
20.5	Реализация ПДВП-анализатора на ПЛИС	448
20.6	Заключение.....	453
20.7	Список использованных источников	454

СПИСОК АВТОРОВ

Петровский Александр Александрович,

д.т.н., профессор, заведующий кафедрой электронных вычислительных средств, научный руководитель НИЛ 3.1. «Мультипроцессорные системы реального времени» БГУИР

Азаров Илья Сергеевич,

к.т.н., доцент кафедры электронных вычислительных средств БГУИР

Борович Адам,

доктор, доцент факультета информатики Белостокского политехнического института

Вашкевич Максим Иосифович,

аспирант кафедры электронных вычислительных средств БГУИР

Качинский Михаил Вячеславович,

к.т.н., доцент кафедры электронных вычислительных средств БГУИР

Клюс Владимир Борисович,

к.т.н., доцент кафедры электронных вычислительных средств БГУИР

Лившиц Михаил Зенадьевич,

к.т.н., доцент кафедры электронных вычислительных средств БГУИР

Лихачев Денис Сергеевич,

к.т.н., доцент кафедры электронных вычислительных средств БГУИР

Павловец Александр Николаевич,

аспирант-заочник кафедры электронных вычислительных средств БГУИР

Парфенюк Марек,

доктор, доцент факультета информатики Белостокского политехнического института

Петровский Алексей Александрович,

к.т.н., доцент кафедры электронных вычислительных машин БГУИР

Родионов Максим Михайлович,

ассистент кафедры электронных вычислительных средств БГУИР

Станкевич Андрей Владимирович,

к.т.н., доцент кафедры электронных вычислительных средств БГУИР

ПРЕДИСЛОВИЕ

Термин «мультимедиа» в настоящее время приобрел необычную популярность. Это объясняется тем, что появилась возможность связывания в единое целое визуальной, акустической и текстуальной информации, т.е. сделан большой шаг от простых массовых продуктов, таких как сотовый телефон и MP3 проигрыватель к амбициозным проектам - человеко-машинным интерфейсам. Стремительное внедрение цифровых технологий в системы радиовещания и телевидения, непрерывное увеличение передач через Интернет обуславливает поиск новых методов и алгоритмов построения анализаторов речевых и звуковых сигналов. В данной книге обобщен многолетний опыт сотрудников НИЛ 3.1 «Мультипроцессорные системы реального времени» БГУИР проектирования различных компонентов систем мультимедиа таких как субполосное кодирование аудио- и речевых сигналов, подавление шумов окружающей среды в речевых сигналах, компенсация акустического и электрического эхо-сигналов, и т.д. Главные направления исследований в данной области заключаются в получении новых функциональных свойств, повышении качества алгоритмов обработки, а также в увеличении эффективности систем мультимедиа. Параллельно с теоретическими результатами, показаны практические реализации на базе современных технологий таких, как реконфигурируемые и программируемые схемы, цифровые процессоры обработки сигналов.

В последнее время все большее число студентов и инженеров интересуются субполосной обработкой информации. Обычно они уже имеют необходимые знания в цифровой обработке сигналов, но теория анализа и синтеза банков фильтров иногда трудна в понимании, особенно, если она представлена в формальной, не визуальной форме. Для выхода из данной ситуации предлагается, в определенной степени, замещение математических выкладок или их дублирование MATLAB кодами, которые более «дружелюбны» для людей знакомых с программированием. Более того, читатель поощряется к выполнению его собственных экспериментов – многочисленными примерами проектирования и практическими результатами, которые помогут ему. Описание полных систем мультимедиа может служить толчком для его более профессиональных устремлений.

Материал книги сгруппирован в три части: методы субполосного анализа; алгоритмы анализа и кодирования речевых и звуковых сигналов в системах мультимедиа; аппаратная реализация анализаторов речевых и звуковых сигналов.

В **первой главе** приводится краткое введение в банки цифровых фильтров, даются основные определения и понятия: полное восстановление или перфективная реконструкция, параунитарный банк, субполосное кодирование, скорость передачи. Рассматривается общий случай схемы субполосного кодера. Выводятся оценки ошибки реконструкции сигнала, а также оптимальное распределение бит по каналам. Доказывается эффективность субполосного кодирования по отношению к полнополосному кодированию.

В данной главе (**глава 2**) с практической точки зрения рассматриваются важные концепции из области многоскоростной обработки сигналов. Вводятся понятия дедимации — понижения частоты дискретизации и интерполяции — повышения частоты дискретизации, а так-

же полифазного представления КИХ-фильтров. Кроме того, показывается, что развитие базовых концепций многоскоростной обработки сигналов приводит к построению эффективных структур банков фильтров. Механизм получения полифазных банков фильтров рассматривается на примере ДПФ-модулированного банка фильтров, показано, как можно получить неравнополосный ДПФ-модулированный банк фильтров, используя фазовое преобразование.

Параунитарные банки фильтров (ПУБФ) (**глава 3**) могут рассматриваться как наиболее значимое преобразование среди многоскоростных систем цифровой обработки сигналов. Это обусловлено тем фактом, что подобные банки являются преобразованиями без потерь в дополнение к гарантированной перфективной реконструкции сигнала. Серьезной практической проблемой, связанной с реализацией ПУБФ на арифметике с фиксированной запятой, является потеря свойства перфективной реконструкции сигнала банком фильтров. Это обусловлено чувствительностью структуры ПУБФ к квантованию коэффициентов. В данной главе представлен новый подход к синтезу и конструированию решетчатых структур 4-х и 8-ми канальных ПУБФ на основе алгебры кватернионов. В отличие от известных решений, предложенные структуры являются структурами без потерь даже при выполнении вычислений с конечной точностью.

Четвертая глава посвящена дискретному вейвлет преобразованию. В ней описывается преобразования, построение дерева пакета дискретного вейвлет преобразования (ПДВП), а также показаны алгоритмы обработки звуковых сигналов с адаптивным частотно-временным планом на основе ПДВП.

В контексте перцептуальной обработки речевых сигналов в **главе 5** рассматривается ДПФ с неравномерным частотным разрешением (от англ. Warped discrete Fourier transform (WDFT)). Коэффициенты WDFT преобразования равномерно распределены в критических частотных полосах, поэтому в психоакустической модели на базе WDFT могут быть уравновешены как хорошее частотное, так и временное разрешение. Разработан метод минимизации ошибки реконструкции сигнала (вычисления обратного WDFT). Предложен алгоритм построения данного преобразования с избыточным базисом, обеспечивающим требуемую точность решения задачи синтеза.

Глава 6 содержит описание классических и оригинальных методов, позволяющих выполнять мгновенный гармонический анализ звуковых и речевых сигналов. Рассмотрен способ оценки мгновенных гармонических параметров на основе специального фильтра анализа, который позволяет интегрировать масштабирующую функцию для правильной оценки нестационарных периодических компонент сигнала, также приведены результаты сравнения точности различных методов оценки гармонических параметров применительно к стационарным и нестационарным сигналам.

В **главе 7** предлагается методология построения психоакустических моделей для трех способов субполосной обработки речевых и звуковых сигналов на основе ПДВП, многополосного вейвлет преобразования с использованием гиперкомплексной арифметики – кватернионов, WDFT, когда ошибка аппроксимации критических частотных полос минимальна в области Барков. Приводится процедура расчета маскирующих порогов в пространстве коэффициентов соответствующего преобразования.

В **главе 8** дается определение векторного квантования и описываются его основные принципы. Рассматривается метод векторного квантования с переменной размерностью векторов для параметров гармонической модели речевого сигнала. Особенностью метода является использование закономерностей психоакустики в процедуре квантования амплитуд, что позволяет повысить качество реконструированного речевого сигнала и снизить вычислительную сложность алгоритмов квантования.

В данной **главе 9** рассматривается гибридное описание сигналов, основанное на комбинированном использовании нескольких моделей параметрического представления. Рассматриваются вопросы, связанные с разделением сигнала на гармоническую, шумовую и переходную составляющие. Описываются все необходимые математические и алгоритмические средства, приводятся результаты их реализации.

Описывается система компрессии речевого сигнала на основе синусоидальной модели с антропоморфической обработкой (**глава 10**). Отличительной особенностью данной системы является то, что речь, как на вокализованных, так и на невокализованных участках, представляется в виде суммы синусоидальных компонент. Антропоморфическая обработка заключается в моделировании физиологических процессов в слуховой системе человека и получении информации о степени важности для восприятия человеком частотных составляющих анализируемого сигнала.

В **главе (11)** описывается модель и структура широкополосного перцептуального CELP-кодера с мультиполосным возбуждением и обработкой субполос в порядке убывания их перцептуальной важности. Для управления процессом кодирования используется психоакустическая модель на базе WDFT. Описывается формирование, оптимизация и кодирование структуры банка кодовых книг векторов возбуждения, а также интеграция препроцессора редактирования шума в структуру компрессора.

Описывается метод кодирования речевых сигналов на базе гибридных техник (**глава 12**). Данный подход обусловлен тем фактом, что методы обработки речи в частотной и временной областях обладают своими преимуществами и недостатками. Рассматриваемая здесь гибридная модель использует преимущества декомпозиции речи на периодическую и аperiodическую компоненты для участков вокализованной речи и CELP-модели – для нестационарных (переходных) участков. Анализ речевого сигнала осуществляется с помощью дискретного преобразования Фурье, согласованного с изменением контура частоты основного тона.

Представлена система конверсии голоса (**глава 13**), основанная на модели сепарации речевого сигнала на «гармоники+шум» и переходные фреймы с отдельной конверсией каждой компоненты модели. Неформальные тесты прослушивания показали, что узнаваемость диктора соответствует приблизительно 70%, реконструированная речь характеризуется достаточно высокой разборчивостью.

Описывается новое применение WDFT (**глава 14**) в одноканальных перцептуальных системах редактирования шума в речи и компрессии звуковых сигналов. WDFT используется одновременно как базис для метода спектрального взвешивания так и для психоакустической модели. Обработка сигналов, осуществляемая в критических частотных полосах и более точна в

контексте перцептуального моделирования по сравнению со схемами на базе ДПФ. Субъективные оценки качества реконструированной речи здесь выше.

В **главе 15** рассматриваются вопросы проектирования неравнополосного косиунсно-модулированного банка фильтров для слуховых аппаратов с использованием частотного растяжения и объединения субполос для увеличения точности аппроксимации критических частотных полос. Уделено внимание проектированию фильтра-прототипа, а также выбору коэффициентов децимации субполосных сигналов. Показано несколько подходов к решению задачи выравнивания групповой задержки в банке фильтров.

В настоящей **главе 16** предлагается перцептуальный кодер звука на основе адаптируемого под сигнал и заданный вычислительный ресурс быстрого вейвлет преобразования с динамической трансформацией частотно-временного плана, базирующегося на подходе динамической трансформации алгоритма (ДТА), вычислении перцептуальной энтропии в области вейвлет коэффициентов, эффективном распределении битов кодирования, учитывающего неидеальность преобразования. Основу системы составляет процессор с динамически реконфигурируемой архитектурой.

В данной **главе 17** рассматриваются вопросы реализации алгоритма ПДВП на основе лестничных структур (lifting scheme) при использовании арифметики с фиксированной запятой. В ходе изложения материала освещаются такие вопросы как математический аппарат, позволяющий осуществить переход от алгоритма на основе КИХ фильтрации к алгоритму на базе лестничных структур, а также вопросы реализации на арифметике с фиксированной запятой. В качестве практического результата в завершении главы предлагается архитектура динамически реконфигурируемого поточного процессора ПДВП.

В **главе 18** обсуждаются особенности аппаратной реализации двухпроцессорной вычислительной платформы на основе ЦПОС TMS320C6713 и периферийного процессора, реализуемого на реконфигурируемой логике, для мультимедиа приложений реального времени, представлены характеристики аппаратной платформы. Показана реализация двух систем компрессии речи и звука: комбинированной системы кодирования широкополосных речевых сигналов и перцептуальный кодер звука на базе ПДВП процессора с динамической трансформацией частотно-временного плана.

В **главе 19** рассматривается аппаратный ускоритель к персональному компьютеру (ПК) на базе ПЛИС XC3S400 семейства Spartan III фирмы Xilinx, предназначенный для реализации трудоемких алгоритмов обработки информации. Особое внимание уделено организации обмена с ПК по интерфейсу USB, приводятся результаты исследования и практические рекомендации по повышению пропускной способности при обмене данными между аппаратным ускорителем и ПК по USB интерфейсу.

В **20-ой главе** показана возможность реализации ПДВП-анализатора звуковых сигналов на базе ПЛИС. В основу вычислительного ядра положена MAC-структура. Процессор представляет собой параллельно-поточную структуру для дерева ПДВП, аппроксимирующего критический частотно-временной план.

1 СУБПОЛОСНЫЙ АНАЛИЗ И ЕГО ЭФФЕКТИВНОСТЬ

Петровский Ал.А.

АННОТАЦИЯ

Приводится краткое введение в банки цифровых фильтров, даются основные определения и понятия: полное восстановление или совершенная реконструкция, параунитарный банк, субполосное кодирование, скорость передачи. Рассматривается общий случай схемы субполосного кодера. Выводятся оценки ошибки реконструкции сигнала, а также оптимальное распределение бит по каналам. Доказывается эффективность субполосного кодирования по отношению к полнополосному кодированию. Показано, что если входной сигнал имеет нормальный закон распределения, коэффициенты децимации в каналах равны, банк фильтров – ортогональный, то субполосный кодер обеспечивает равное или лучшее качество по сравнению с полнополосным кодером при любом входном сигнале.

Ключевые слова – банк фильтров, субполосное кодирование, эффективность кодирования, скорость передачи.

СОДЕРЖАНИЕ

1.1	Введение в банки фильтров.....	13
1.1.1	Определения.....	13
1.1.2	Ошибка реконструкции сигнала	17
1.1.3	Скорость передачи субполосного кодера	19
1.2	Оценка эффективности субполосного кодирования.....	20
1.2.1	Оптимальное распределение бит по каналам субполосного кодера.....	20
1.2.2	Дисперсия минимальной ошибки реконструкции сигнала	21
1.2.3	Оценки эффективности субполосного кодирования	21
1.3	Список использованных источников.....	23

1.1 Введение в банки фильтров

1.1.1 Определения

Банк фильтров - цифровая система, состоящая из секций анализа и синтеза, называемых банком фильтров анализа и банком фильтров синтеза (рис. 1.1). Входной сигнал $x(n)$, представленный последовательностью отсчетов, разбивается при помощи фильтров секции анализа $H_k(z)$ ($k = 0, 1, \dots, M - 1$) на M субполосных составляющих, которые в идеальном случае в частотной области не перекрываются. Операции, выполняемые секцией синтеза, являются обратными операциями секции анализа. Подобранным соответствующим образом набор фильтров секции синтеза $F_k(z)$ ($k = 0, 1, \dots, M - 1$), можно восстановить исходный сигнал $y(n)$ из его субполосных компонент [1].

Банк фильтров относится к классу многоскоростных систем цифровой обработки сигналов [1-4], в которых частота дискретизации различна в разных точках системы. Значение коэф-

коэффициента темпа поступления отсчетов в канале (коэффициента децимации m_k) задается утверждением о дискретизации сигнала в зависимости от ширины частотной полосы канала B_k и его положения в банке фильтров. Оригинальная частота дискретизации f_s сигнала в k -ом канале теоретически может быть уменьшена в $m_k \leq f_s/B_k$ раз. Равенство, в данном случае, означает, что канал есть максимально децимированным:

$$\sum_{k=0}^{M-1} \frac{1}{m_k} = 1. \quad (1.1)$$

Таким образом, в частотно-временном представлении сигнала исключена избыточность, т.е. сумма отсчетов во всех каналах соответствует количеству отсчетов в исходном сигнале. Банк фильтров считается передецимированным, если $\sum_{k=0}^{M-1} 1/m_k > 1$, т.е. достаточно, чтобы хотя бы в одном канале коэффициент децимации не был равен числу каналов. Более сложные случаи позиционирования субполос каналов в банках фильтров рассмотрены в [4-6].

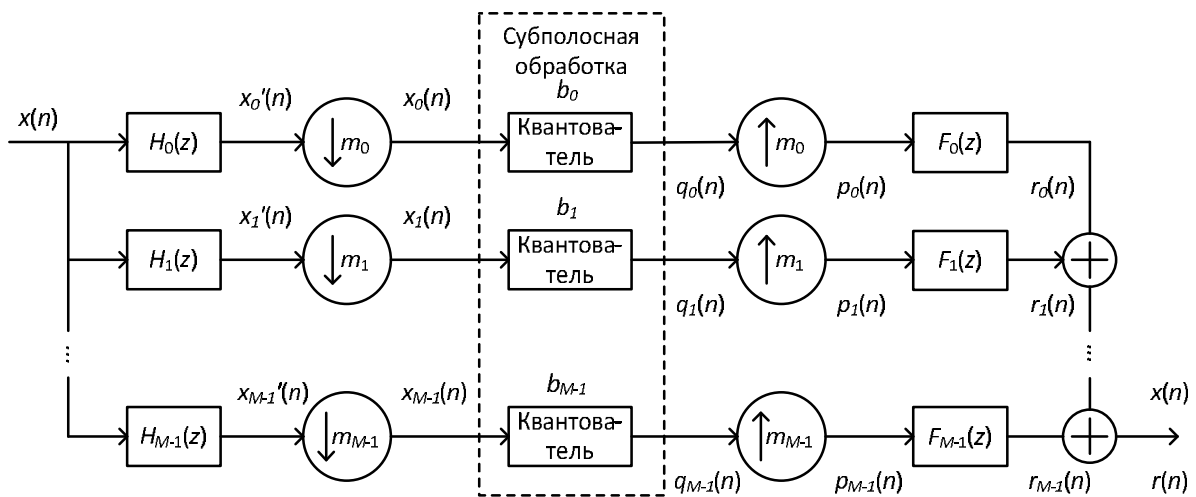


Рис. 1.1 – Банк фильтров: система анализа/синтеза сигнала

Соотношение между z -преобразованием сигналов на входе и выходе M – канального банка фильтров, изображенного на рис. 1.1, является следующим [1]:

$$Y(z) = \sum_{k=0}^{M-1} F_k(z) \frac{1}{m_k} \sum_{l=0}^{m_k-1} H_k(zW_{m_k}^l) X(zW_{m_k}^l), \quad (1.2)$$

где $W_{m_k} = e^{-j2\pi/m_k}$. Анализ данного выражения показывает, что в банке возможны искажения входного сигнала: амплитудные, фазовые и отражения частотных характеристик (элайзинг), обусловленные наличием дециматоров и интерполяторов. Величина данных искажений зависит как от частотных характеристик канальных фильтров, так и выбора коэффициентов децимации m_k . Анализ искажений, возникающих в структуре банка фильтров, проще анализировать в максимально децимированном банке фильтров, для которого соотношение (1.2) значительно упрощается:

$$Y(z) = \sum_{k=0}^{M-1} T_k(z) X(zW_M^k), \quad (1.3)$$

где $T_k(z)$ – передаточная функция k -го канала:

$$T_k(z) = \frac{1}{M} \sum_{l=0}^{M-1} F_l(z) H_l(zW_M^k). \quad (1.4)$$

Выходной сигнал $y(n)$ системы анализа-синтеза банка фильтров будет свободен от элайзинговой составляющей $X(zW_M^k)$, $k > 0$ в случае, если

$$T_k(z) = 0, \text{ для } 1 \leq k \leq M. \quad (1.5)$$

В банке фильтров, для которого справедливо условие (1.5), остаются только амплитудные и фазовые искажения, которые определяются согласно следующему выражению:

$$\frac{Y(z)}{X(z)} = T_0(z) = \frac{1}{M} \sum_{l=0}^{M-1} F_k(z) H_k(z). \quad (1.6)$$

Очевидно, что для получения перфективной реконструкции входного сигнала $x(n)$ банком фильтров, передаточная функция искажений $T_0(z)$ должна принять форму простого звена задержки с некоторым масштабированием амплитуды:

$$T_0(z) = cz^{-\Delta}, c \neq 0, \Delta \in \mathbb{Z}. \quad (1.7)$$

Полное восстановление или перфективная реконструкция – свойство банка цифровых фильтров, заключающееся в том, что сигнал, прошедший через схему анализа-синтеза идентичен входному с точностью до задержки. Для этого фильтры синтеза должны подавлять наложение частотных характеристик (элайзинг) и устранять амплитудные и фазовые искажения [1].

Параунитарный (ортогональный) банк фильтров (ПУБФ) – банк фильтров, у которого передаточные функции анализирующих и синтезирующих фильтров и их соответственно смещенные версии ортогональны друг другу. Фильтры синтеза в параунитарных банках являются транспонированными версиями фильтров анализа [1]:

$$F_k(z) = H_k^T(z^{-1}). \quad (1.8)$$

При соблюдении этого условия обеспечивается возможность перфективной реконструкции банком фильтров входного сигнала $x(n)$: пусть $x[n] = [x_0[n] \dots x_{M-1}[n]]^T$ будет входным вектором и $y[n] = [y_0[n] \dots y_{M-1}[n]]^T$ будет соответствующим выходным вектором с $M \times M$ параунитарной передаточной матрицей $A(z)$. Пусть $S_{xx}(e^{j\omega})$ будет $M \times M$ СПМ-матрица входного вектора $x[n]$. Заметим, что СПМ i -го входного компонента $x_i[n]$ является i -й элемент $S_{xx}(e^{j\omega})$. Поэтому дисперсия $x_i[n]$ составит

$$\int_{-\pi}^{\pi} (S_{xx}(e^{j\omega}))_i \frac{d\omega}{2\pi}. \quad (1.9)$$

Усредненная дисперсия входного сигнала будет

$$\frac{1}{N} \sum_{i=0}^{N-1} \sigma_i^2 = \frac{1}{N} \int_{-\pi}^{\pi} \text{tr} (S_{xx}(e^{j\omega})) \frac{d\omega}{2\pi}, \quad (1.10)$$

где $tr(B)$ – след матрицы B .

СПМ вектора на выходе определяется как

$$S_{yy}(e^{jw}) = A(e^{jw})S_{xx}(e^{jw})A^H(e^{jw}), \quad (1.11)$$

где A^H - матрица, эрмитово транспонированная к матрице A .

Усредненная дисперсия выхода составит

$$\frac{1}{N} \int_{-\pi}^{\pi} tr(S_{yy}(e^{jw})) \frac{dw}{2\pi} = \frac{1}{N} \int_{-\pi}^{\pi} tr(A(e^{jw})S_{xx}(e^{jw})A^H(e^{jw})) \frac{dw}{2\pi}. \quad (1.12)$$

Так как $tr(AB) = tr(BA)$, это упрощает выражение (1.12)

$$\frac{1}{N} \int_{-\pi}^{\pi} tr(S_{xx}(e^{jw})A^H(e^{jw})A(e^{jw})) \frac{dw}{2\pi}.$$

Так как матрица $A(z)$ - параунитарная, $A^H(e^{jw})A(e^{jw}) = I$, следовательно

$$\frac{1}{N} \int_{-\pi}^{\pi} tr(S_{xx}(e^{jw})) \frac{dw}{2\pi} \quad (1.13)$$

или усредненная дисперсия выхода равна усредненной дисперсии входа. Этот факт показывает, что в параунитарной системе энергия сохраняется.

Банк фильтров можно представить в полифазной форме, если передаточные функции секций анализа и синтеза записать в виде соответствующих векторов:

$$\begin{aligned} \mathbf{H}(z) &= [H_0(z) H_1(z) \dots H_{M-1}(z)]^T, \\ \mathbf{F}(z) &= [F_0(z) F_1(z) \dots F_{M-1}(z)]^T, \end{aligned} \quad (1.14)$$

то тогда можно выбрать такие полифазные матрицы анализа

$$\mathbf{E}(z) = \begin{bmatrix} E_{0,0}(z) & \dots & E_{0,M-1}(z) \\ \vdots & \ddots & \vdots \\ E_{M-1,0}(z) & \dots & E_{M-1,M-1}(z) \end{bmatrix}, \quad (1.15)$$

и синтеза

$$\mathbf{D}(z) = \begin{bmatrix} D_{0,0}(z) & \dots & D_{0,M-1}(z) \\ \vdots & \ddots & \vdots \\ D_{M-1,0}(z) & \dots & D_{M-1,M-1}(z) \end{bmatrix}, \quad (1.16)$$

что вектора передаточных функций секций анализа и синтеза можно представить следующим образом [1]:

$$\begin{aligned} \mathbf{H}(z) &= \mathbf{E}(z^M) [1 \ z^1 \ \dots \ z^{-(M-1)}]^T, \\ \mathbf{F}(z) &= [z^{-(M-1)} \ z^{-(M-2)} \ \dots \ 1]^T \mathbf{D}(z^M). \end{aligned} \quad (1.17)$$

Для получения перфективной реконструкции на компоненты полифазных матриц накладывается дополнительное ограничение:

$$\mathbf{D}(z)\mathbf{E}(z) = cz^{-\Delta}\mathbf{I}, \quad c \neq 0, \Delta \geq 0, \quad (1.18)$$

где c – ненулевая константа; Δ –задержка, выраженная целым числом интервалов дискретизации, вносимая секциями анализа-синтеза; \mathbf{I} – единичная матрица. На рис. 1.2 показана полифазная структура банка фильтров.

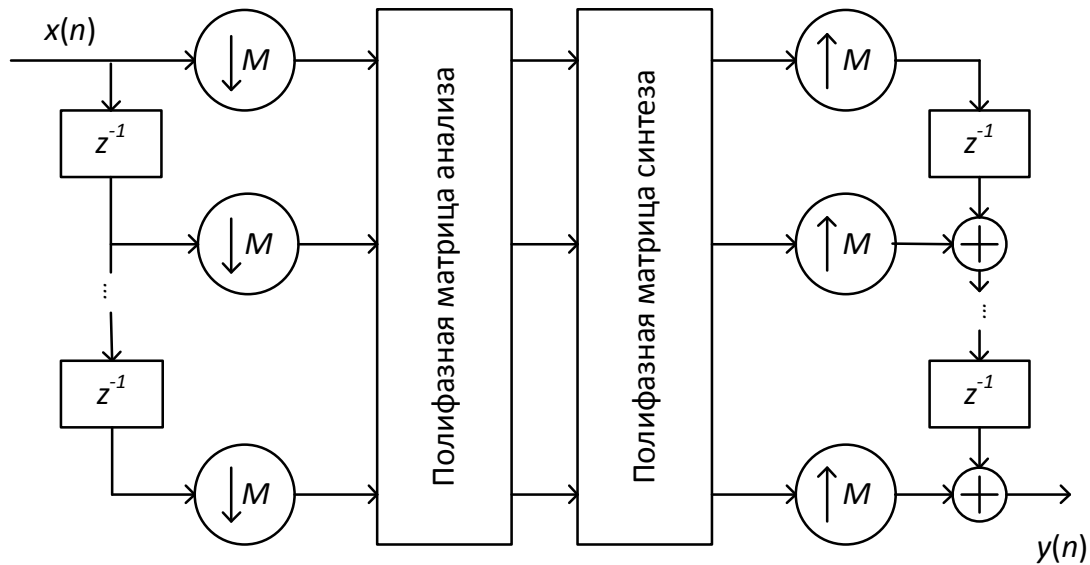


Рис. 1.2 – Полифазная структура банка фильтров

Традиционно БФ разделяют на банки с равнополосными и неравнополосными каналами, ортогональные, биортогональные, двухканальные и многоканальные и т.д. Каждый фильтр банка цифровых фильтров образует канал. Поэтому говорят об M -канальном банке фильтров. Сигнал в канале называется субполосой, отсюда название субполосная фильтрация или субполосное кодирование [3,4].

1.1.2 Ошибка реконструкции сигнала

Рассматривается субполосный кодер, представленный на рис. 1.1, состоящий из $k = 0, \dots, M - 1$ каналов. Каждый канал имеет анализирующий фильтр $H_k(z)$, синтезирующий фильтр $F_k(z)$, дециматор/интерполятор с коэффициентом передискретизации m_k и b_k -битный квантователь. Входной сигнал $x(n)$ после фильтрации анализирующими фильтрами на рис. 1.1 обозначен как $x'(n)$, а каналные сигналы - $x_k(n)$. Субполосный кодер максимально децимирован, если выполняется условие (1.1).

Пусть входной сигнал $x(n)$ является стационарным в широком смысле и имеет среднее значение, равное нулю. Следовательно, все последующие (производные) сигналы (включая шум квантования) также будут иметь среднее значение, равное нулю. Это предположение не является причиной возникновения каких-либо трудностей для применения субполосного кодирования к сигналам со средним значением, отличным от нуля, таких как изображения, так как система в этом случае эквивалентна системе с нулевым средним, полученной путем вычитания среднего значения из входного сигнала [7].

Следовательно, если $S_{xx}(e^{j\omega})$ является спектральной плотностью мощности (СПМ) сигнала $x(n)$, то его дисперсия определяется как [1]

$$\sigma_x^2 = \int_{-\pi}^{\pi} S_{xx}(e^{j\omega}) \frac{\partial \omega}{2\pi}.$$

Дисперсия сигнала $x'_k(n)$ на выходе анализирующего фильтра $H_k(z)$, будет определяться следующим выражением [1]

$$\sigma_k^2 = \int_{-\pi}^{\pi} S_{xx}(e^{j\omega}) |H_k(e^{j\omega})|^2 \frac{\partial \omega}{2\pi},$$

для $k = 0, 1, \dots, M - 1$. Так как сигнал стационарен в широком смысле, то дисперсия не изменится после децимации [1,7], т.е. дисперсия $x_k(n)$ субполосы k равна σ_k^2 . Квантователь представляется как модель с аддитивным шумом [8], т.е. выходом квантователя является сигнал $x_k(n) + q_k(n)$, где $x_k(n)$ - входной сигнал и $q_k(n)$ - шум квантования. Дисперсия шума квантования, как для равномерных квантователей, так и для квантователей, адаптированных под сигнал, определяется как [1,8]

$$\sigma_{q_k}^2 = \varepsilon_k 2^{-2b_k} \sigma_k^2, \quad (1.19)$$

где ε_k – константа, характеризующая квантователь, которая зависит от функции плотности вероятности k -го субполосного сигнала.

Предположим, используются равномерные (многобитные) квантователи, тогда шумы квантования $q_k(n)$ различных каналов некоррелированы между собой и являются «белым» шумом [9]. СПМ шума $q_k(n)$ является константой $\sigma_{q_k}^2$, так как это «белый» шум, и имеют нулевое среднее. Сигналы $p_k(n)$, полученные после интерполяции $q_k(n)$, определяются как

$$p_k(n) = \begin{cases} q_k(n/m_k), & \text{если } n \bmod m_k = 0 \\ 0 & \text{иначе} \end{cases},$$

и больше не являются стационарными в широком смысле, но имеют циклическую стационарность с периодом m_k [7]. Другими словами, если $n \bmod m_k = 0$, то СПМ сигнала $p_k(n)$ равна $\sigma_{q_k}^2$, а СПМ оставшихся отсчетов равна нулю. Отсчеты шума $p_k(n)$ поступают на вход синтезирующего фильтра $F_k(n)$, на выходе которого получается сигнал $r_k(n)$, стационарный в широком смысле с периодом m_k . Усредненная на периоде m_k дисперсия сигнала $r_k(n)$ определяется как

$$\frac{1}{m_k} \left[\int_{-\pi}^{\pi} \sigma_{q_k}^2 |F_k(e^{j\omega})|^2 \frac{\partial \omega}{2\pi} + 0 + \dots + 0 \right],$$

где $m_k - 1$ – нули, полученные из нулевых отсчетов сигнала шума $p_k(n)$. Ошибка реконструкции $r(n)$ равна сумме всех ошибок $r_k(n)$, что обусловлено линейностью секции синтеза банка фильтров. Так как $r_k(n)$ во всех каналах некоррелированы, то дисперсия их суммы равна сумме их дисперсий [10]

$$\sigma_r^2 = \sum_{k=0}^{M-1} \frac{\sigma_{q_k}^2}{m_k} \int_{-\pi}^{\pi} |F_k(e^{j\omega})|^2 \frac{\partial \omega}{2\pi}, \quad (1.20)$$

где σ_r^2 - дисперсия ошибки реконструкции. Пусть $n_k = \int_{-\pi}^{\pi} |F_k(e^{j\omega})|^2 \frac{\partial \omega}{2\pi}$ означает нормы синтезирующих фильтров (заметим, что для КИХ-фильтра с коэффициентами импульсной характеристики (h_0, h_1, \dots, h_L) норма равна $h_0^2 + h_1^2 + \dots + h_L^2$). Затем, используя уравнение (1.19), выражение (1.20) преобразуется к следующему виду

$$\sigma_r^2 = \sum_{k=0}^{M-1} \frac{\varepsilon_k 2^{-2b_k} \sigma_k^2 n_k}{m_k}. \quad (1.21)$$

Для параунитарных систем [1] свойство отсутствия потерь (или энергетический баланс) говорит о том, что энергия выхода равна энергии входа (1.13). Вследствие этого факта, дисперсия ошибки реконструкции сигнала банком фильтров определяется следующим выражением

$$\sigma_r^2 = \sum_{k=0}^{M-1} \frac{\varepsilon_k 2^{-2b_k} \sigma_k^2}{m_k}. \quad (1.22)$$

Сравнивая уравнения (1.21) и (1.22) для ортогонального случая можем заметить, что в (1.22) отсутствует норма синтезирующего фильтра n_k . Это является следствием свойства отсутствия потерь, подразумевающего фильтры с единичной энергией или $n_k = 1$ для всех k [1]. Следует отметить также, что допущение об ортогональном банке фильтров заменяет допущение о равномерных (многобитных) квантователях и не является необходимым для получения данного результата. Поэтому, в смысле квантователя, результат более общий для ортогонального случая.

1.1.3 Скорость передачи субполосного кодера

Для субполосного кодера на рис. 1.1 b_k определяет количество бит на отсчет для k -го квантователя. Однако, благодаря коэффициенту децимации m_k , данным квантователем квантуется один отсчет для каждого из m_k входных отсчетов. Поэтому скорость передачи квантователя равна b_k/m_k бит на входной отсчет. Следовательно, усредненная скорость передачи субполосной системы можно определить как

$$b = \sum_{k=0}^{M-1} \frac{b_k}{m_k} \text{ бит/отсчет}. \quad (1.23)$$

Полнополосный кодер просто квантует вход $x(n)$, используя b -битный квантователь. Поэтому, из уравнения (1.22), его дисперсия ошибки квантования (которая также является дисперсией ошибки реконструкции) определяется как

$$\sigma_q^2 = \varepsilon 2^{-2b} \sigma_x^2, \quad (1.24)$$

где ε определяется как и ранее.

Эффективность субполосного кодирования вычисляется как отношение дисперсии ошибки реконструкции полнополосного кодера (или импульсно-кодовой модуляции – ИКМ) σ_q^2 к дисперсии ошибки реконструкции субполосного кодера σ_r^2 с аналогичной скоростью передачи данных [8].

1.2 Оценка эффективности субполосного кодирования

1.2.1 Оптимальное распределение бит по каналам субполосного кодера

Проблема оптимального распределения бит в каналах состоит в нахождении b_0, \dots, b_{M-1} , которые минимизируют дисперсию ошибки реконструкции σ_r^2 в уравнении (1.21), удовлетворяя ограничению (1.23). Данная минимизация с ограничением (1.23) может быть решена с использованием метода множителей Лагранжа. Для этого определяется целевая функция

$$C = \sum_{k=0}^{M-1} \frac{\varepsilon_k 2^{-2b_k} \sigma_k^2 n_k}{m_k} + \lambda \left(\sum_{k=0}^{M-1} \frac{b_k}{m_k} - b \right),$$

где λ - множитель Лагранжа. Дифференцируя C по b_k и приравнявая к нулю, ($\partial C / \partial b_k = 0$) или

$$\frac{1}{m_k} \varepsilon_k 2^{-2b_k} \sigma_k^2 n_k (-2 \ln 2) + \lambda \frac{1}{m_k} = 0,$$

из которого следует, что

$$\frac{\partial}{\partial b_k} 2^{-2b_k} = \frac{\partial}{\partial b_k} e^{-2 \ln 2 b_k} = -2 \ln 2 e^{-2 \ln 2 b_k}.$$

Таким образом,

$$2 \ln 2 \varepsilon_k 2^{-2b_k} \sigma_k^2 n_k = \lambda,$$

или

$$2^{-2b_k} = \frac{\lambda}{2 \ln 2 \varepsilon_k \sigma_k^2 n_k}. \quad (1.25)$$

После логарифмирования по основанию 2 обеих частей (1.25) получается, что

$$-2b_k = \log_2 \frac{\lambda}{2 \ln 2 \varepsilon_k \sigma_k^2 n_k},$$

или

$$b_k = \frac{1}{2} \log_2 \frac{2 \ln 2 \varepsilon_k \sigma_k^2 n_k}{\lambda} = \frac{1}{2} \log_2 \frac{2 \ln 2}{\lambda} + \frac{1}{2} \log_2 (\varepsilon_k \sigma_k^2 n_k). \quad (1.26)$$

Данное выражение действительно для всех значений $k = 0, \dots, M - 1$. Подставляя уравнение (1.26) в ограничение (1.23), получается, что

$$\begin{aligned} b &= \sum_{k=0}^{M-1} \frac{1}{m_k} \left[\frac{1}{2} \log_2 \frac{2 \ln 2}{\lambda} + \frac{1}{2} \log_2 (\varepsilon_k \sigma_k^2 n_k) \right] \\ &= \frac{1}{2} \log_2 \frac{2 \ln 2}{\lambda} \sum_{k=0}^{M-1} \frac{1}{m_k} + \frac{1}{2} \sum_{k=0}^{M-1} \log_2 (\varepsilon_k \sigma_k^2 n_k) \frac{1}{m_k}. \end{aligned} \quad (1.27)$$

Первая сумма здесь равна 1 на основании уравнения (1.16). Вторая сумма логарифмов может быть записана как логарифм произведения. Используя эти упрощения, средняя скорость передачи в субполосной системе равна

$$b = \frac{1}{2} \log_2 \frac{2 \ln 2}{\lambda} + \frac{1}{2} \log_2 \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}}, \quad (1.28)$$

или

$$\frac{1}{2} \log_2 \frac{2 \ln 2}{\lambda} = b - \frac{1}{2} \log_2 \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}}. \quad (1.29)$$

Подставляя данный результат в уравнение (1.26), получается следующее распределение бит по каналам:

$$b_k = b - \frac{1}{2} \log_2 \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}} + \frac{1}{2} \log_2 (\varepsilon_k \sigma_k^2 n_k) = b + \frac{1}{2} \log_2 \frac{\varepsilon_k \sigma_k^2 n_k}{\prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}}}, \quad (1.30)$$

для всех значений $k = 0, \dots, M - 1$.

Это и есть оптимальное распределение бит по каналам субполосного кодера. Заметим, что для ортогонального случая оптимальное распределение бит получается путем подстановки $n_k = 1$ для всех k . Следует отметить, что результирующая скорость передачи – действительная величина и может быть отрицательной.

1.2.2 Дисперсия минимальной ошибки реконструкции сигнала

На основании оптимального распределения бит (уравнение (1.30)) следует, что

$$2^{-2b_k} = 2^{-2b} \frac{\prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}}}{\varepsilon_k \sigma_k^2 n_k},$$

для всех $k = 0, \dots, M - 1$. Подставляя это выражение в уравнение (1.22) и принимая во внимание (1.1), получается, что дисперсия минимальной ошибки реконструкции сигнала равна

$$\sigma_r^2 = \sum_{k=0}^{M-1} \frac{1}{m_k} 2^{-2b} \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}} = 2^{-2b} \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}} \sum_{k=0}^{M-1} \frac{1}{m_k} = 2^{-2b} \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}}. \quad (1.31)$$

1.2.3 Оценки эффективности субполосного кодирования

Эффективность субполосного кодирования можно определить как отношение дисперсии ошибки реконструкции полнополосного кодера σ_q^2 (1.24) к дисперсии ошибки реконструкции субполосного кодера σ_r^2 (1.31) с аналогичной скоростью передачи данных

$$G = \frac{\varepsilon 2^{-2b} \sigma_x^2}{2^{-2b} \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}}} = \frac{\varepsilon \sigma_x^2}{\prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{\frac{1}{m_i}}}. \quad (1.32)$$

Некоторые специальные случаи оценки эффективности субполосного кодирования рассматриваются ниже.

Предположим входной сигнал $x(n)$ является гауссовым (отсчеты имеют гауссову функцию плотности вероятности). Тогда известно, что выход линейной системы с гауссовым входом также является гауссовым [10]. Таким образом, все субполосные сигналы будут гауссовыми, что приведет к $\varepsilon = \varepsilon_0 = \varepsilon_1 = \dots = \varepsilon_{M-1}$ (при условии многобитного квантователя). Следовательно, эффективность кодирования составит

$$G = \frac{\varepsilon \sigma_x^2}{\left(\varepsilon^{\frac{1}{m_0} + \dots + \frac{1}{m_{M-1}}} \prod_{i=0}^{M-1} (\sigma_i^2 n_i)^{\frac{1}{m_i}} \right)} = \frac{\sigma_x^2}{\prod_{i=0}^{M-1} (\sigma_i^2 n_i)^{\frac{1}{m_i}}}.$$

Пусть коэффициенты децимации будут равны $m_0 = m_1 = \dots = m_{M-1}$. Из уравнения (1.1) следует, что каждый коэффициент равен M и эффективность субполосного кодирования составит

$$G = \frac{\varepsilon \sigma_x^2}{[\prod_{i=0}^{M-1} (\sigma_i^2 n_i)]^{\frac{1}{M}}}.$$

В случае ортогонального банка фильтров $n_k = 1$, как было обозначено ранее, эффективность субполосного кодера будет определяться как

$$G = \frac{\varepsilon \sigma_x^2}{\prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2)^{\frac{1}{m_i}}}. \quad (1.33)$$

Далее, свойство отсутствия потерь, обозначенное ранее, может быть применено к матрице полифазного анализа (1.15). Так как входом полифазного анализа является $x(n)$, то усредненная дисперсия входа будет равна σ_x^2 . Пусть σ_k^2 обозначает дисперсию выхода $k = 0, \dots, M - 1$ субполос, тогда на основании свойства отсутствия потерь следует, что

$$\sigma_x^2 = \sum_{k=0}^{M-1} \frac{\sigma_k^2}{m_k}. \quad (1.34)$$

Подставляя равенство (1.34) в соотношение (1.33) следует, что эффективность субполосного кодирования может быть выражена следующим образом:

$$G = \frac{\varepsilon \sum_{i=0}^{M-1} \frac{\sigma_i^2}{m_i}}{\prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2)^{\frac{1}{m_i}}}. \quad (1.35)$$

Анализ (1.35) показывает, что более простое выражение оценки эффективности субполосного кодирования получается при одновременном рассмотрении следующих допущений [11-15]: входной сигнал имеет нормальный закон распределения, коэффициенты децимации в каналах равны, банк фильтров - ортогональный. При этом, эффективность субполосного кодирования определяется отношением среднего арифметического к среднему геометрическому неотрицательных величин σ_i^2 :

$$G = \frac{\frac{1}{M} \sum_{i=0}^{M-1} \sigma_i^2}{\left(\prod_{i=0}^{M-1} \sigma_i^2\right)^{\frac{1}{M}}} \quad (1.36)$$

Так как среднее арифметическое больше или равно среднему геометрическому, следовательно эффективность субполосного кодирования $G \geq 1$, т.е. субполосный кодер обеспечивает равное или лучшее качество по сравнению с полнополосным кодером при любом входном сигнале. Величина обратная выражению (1.36) также известно как мера пологости спектра (*SFM*) [16,17]. Часто, в задачах кодирования речи используется взвешенная на порог маскирования величина *SFM* (перцептуально взвешенная *SFM* – *PSFM*) [17,18].

1.3 Список использованных источников

1. Vaidynathan P.P., Multirate systems and filter banks, Prentice Hall: Englewood Cliffs, 1993.
2. Crochiere R.E., Rabiner L., Multirate digital signal processing, Prentice-Hall, Englewood Cliffs, NJ, USA, 1983.
3. Витязев В.В., Цифровая частотная селекция сигналов, Радио и связь, М., 1993, - 240с.
4. Piotrowski A., Parfieniuk M., Cyfrowe banki filtrow: analiza, synteza i implementacja dla systemow multimedialnych, Politechnika Bialostocka, Bialystok, 2006, - 389p.
5. Vaughan R.G., Scott N.L., White D.R., The theory of bandpass sampling, IEEE Trans. Signal processing, vol. 39, no. 9, pp. 1973-1984, 1991.
6. Parfieniuk M., Petrovsky A., Simple rule of selection of subsampling ratios for warped filter banks, in Proc. VIII Int. conf. "Modern communication systems", Naroch, Belarus, 2003, pp. 130-134, Special Issue of Trans. Belarussian Engineer Academy, no. 1(15)/3.
7. Sathe V.P., Vaidynathan P.P., Effects of multirate systems on the statistical properties of random signals, IEEE Transactions Signal Processing, 41(1), January 1993, pp. 131-146.
8. Jaynt N.S., Noll P., Digital coding of waveforms, Prentice Hall: Englewood Cliffs, 1984
9. Uzun N., Haddad R.A., Cyclostationary modeling, analysis, and optimal compensation of quantization errors in subband codecs, IEEE Transactions Signal Processing, 43(9), September 1995, pp/ 2109-2119.
10. Papoulis A., Probability, random variables, and stochastic processes, McGraw-Hill: Tokyo, 1984.
11. Soman A.K., Vaidynathan P.P., Coding gain in paraunitary analysis/synthesis systems, IEEE Transactions Signal Processing, 41(5), May 1993, pp. 1824-1835.
12. Djokovic I., Vaidynathan P.P., On optimal analysis/synthesis filters for coding gain maximization, IEEE Transactions Signal Processing, 44(5), May 1996, pp. 1276-1279.
13. Calvagno G., Mian G.A., Rinaldo R., Computation of the coding gain for subband coders, IEEE Transactions Communication, 44(4), April 1996, pp. 475-487.

-
14. Kok C.W., Nguyen T.Q., Multirate filter banks and transform coding gain, IEEE Transactions Signal Processing, 46(7), July 1998, pp. 2041-2044.
 15. Gosse K., Duhamel P., Perfect reconstruction versus MMSE filter banks in source coding, IEEE Transactions Signal Processing, 45(9), September 1997, pp. 2188-2202.
 16. Spanias A., Painter T., Atti V., Audio signal processing and coding, Wiley-Interscience, NJ, USA, 2007.
 17. Bosi M., Goldberg R.E., Introduction to digital audio coding and standards, Springer Science+Business Media, USA, 2003.
 18. Петровский А.А., Бежавский К., Петровский Ал.А., Перцептуальное кодирование аудио и речевых сигналов, Доклады БГУИР, № 1(5), 2004, с. 73-91.

2 ВВЕДЕНИЕ В ЦИФРОВЫЕ БАНКИ ФИЛЬТРОВ

Вашкевич М.И., Парфенюк М., Петровский А.А.

АННОТАЦИЯ

В данной главе с практической точки зрения рассматриваются важные концепции из области многоскоростной обработки сигналов. Вводятся понятия децимации — понижения частоты дискретизации и интерполяции — повышения частоты дискретизации, а также полифазного представления КИХ-фильтров. Кроме того, показывается, что развитие базовых концепций многоскоростной обработки сигналов приводит к построению эффективных структур банков фильтров. Механизм получения полифазных банков фильтров рассматривается на примере ДПФ-модулированного банка фильтров. В последнем параграфе показано, как используя фазовое преобразование можно получить неравнополосный ДПФ-модулированный банк фильтров.

Работа частично финансировалась по гранту W/WI/06/09 Белостокским политехническим институтом.

Ключевые слова – децимация, интерполяция, полифазное представление фильтра, банк фильтров с почти полными восстановлением.

СОДЕРЖАНИЕ

2.1	Введение.....	25
2.2	Изменение частоты дискретизации	26
2.2.1	Децимация	26
2.2.2	Интерполяция	27
2.2.3	Эквивалентные схемы включения интерполяторов и дециматоров.....	28
2.3	Построение банков фильтров на основе модуляции.....	29
2.3.1	Основные понятия	29
2.3.2	ДПФ-модулированный банк фильтров	31
2.3.3	Неравнополосный ДПФ-модулированный банк фильтров.....	40
2.3.4	Полифазная структура неравнополосного ДПФ-модулированного банка фильтров.....	41
2.3.5	Примеры проектирования неравнополосных ДПФ-модулированных банков фильтров.....	45
2.4	Список использованных источников.....	48

2.1 Введение

Растущая потребность в современных цифровых системах обработки данных с более чем одной частотой дискретизации привела к развитию нового направления цифровой обработки сигналов (ЦОС), известного как многоскоростная обработка сигналов либо обработка сигналов при нескольких скоростях (от англ. multirate signal processing) [1]. В подобной обработке сигналов основную роль играют операции децимации и интерполяции. Децимация уменьшает частоту дискретизации, эффективно сжимая данные и оставляя только необходимую информацию. Интерполяция, наоборот, увеличивает частоту дискретизации. Дальнейшее развитие этих принципов ведет к построению банков фильтров — устройств, выполняющих

разбиение входного сигнала на заданное число субполосных сигналов. Каждый такой сигнал содержит в себе частоты только определенного диапазона, вследствие чего можно понизить его частоту дискретизации. Подобные способы обработки сигналов, включающие в себя понижение частоты дискретизации с последующей субполосной обработкой и восстановлением сигнала, получили большое распространение в таких областях как кодирование речи, спектральный анализ, кодирование изображений, редактирование шумов, повышение качества речевого сигнала и другие.

2.2 Изменение частоты дискретизации

2.2.1 Децимация

Дециматор — стандартный блок, используемый в многоскоростной обработке сигналов для снижения частоты дискретизации в M раз. Графическое изображение операции децимации показано на рис. 2.1. Работа дециматора описывается следующим уравнением

$$y_D(n) = x(Mn),$$

которое говорит о том, что каждое выходное значение в момент времени n равно входному значению в момент времени Mn . Таким образом, из входных отсчетов на выход проходят только отсчеты с номерами кратными M [2]. Уменьшение скорости следования отсчетов при децимации для случая $M = 2$ показано на рис. 2.2.



Рис. 2.1 – Блок-схема дециматора

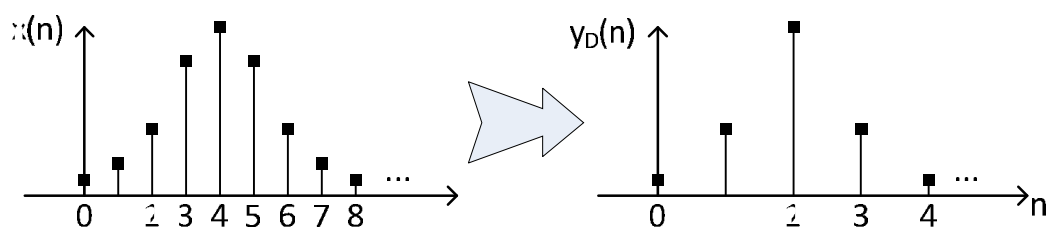


Рис. 2.2 – Пример децимации сигнала для $M = 2$

Z -преобразование выходной последовательности дециматора имеет вид:

$$Y_D(z) = \frac{1}{M} \sum_{k=0}^{M-1} X(z^{1/M} W_M^k), \quad (2.1)$$

где $W_M \triangleq e^{-j2\pi/M}$. В частотной области децимация приводит к тому, что частотный диапазон сигнала сужается в M раз, что может вызвать перекрытие спектров, как показано на рис. 2.3.

Применяя подстановку $z = e^{j\omega}$ к формуле (2.1) получаем выражение для частотной характеристики сигнала после децимации

$$Y_D(e^{j\omega}) = \frac{1}{M} \sum_{k=0}^{M-1} X(e^{j(\omega-2\pi k)/M}), \quad (2.2)$$

из которого становится понятным появление копий спектра на рис. 2.3, сдвинутых на $2\pi/M$. Для того, чтобы избежать подобных наложений необходимо до операции децимации пропустить сигнал через цифровой фильтр нижних частот – фильтр защиты от наложения спектров. Фильтр должен быть спроектирован таким образом, чтобы выходной сигнал имел ограниченный частотный диапазон $-\pi/M < \omega < \pi/M$, что воспрепятствует наложению спектров в частотной области.

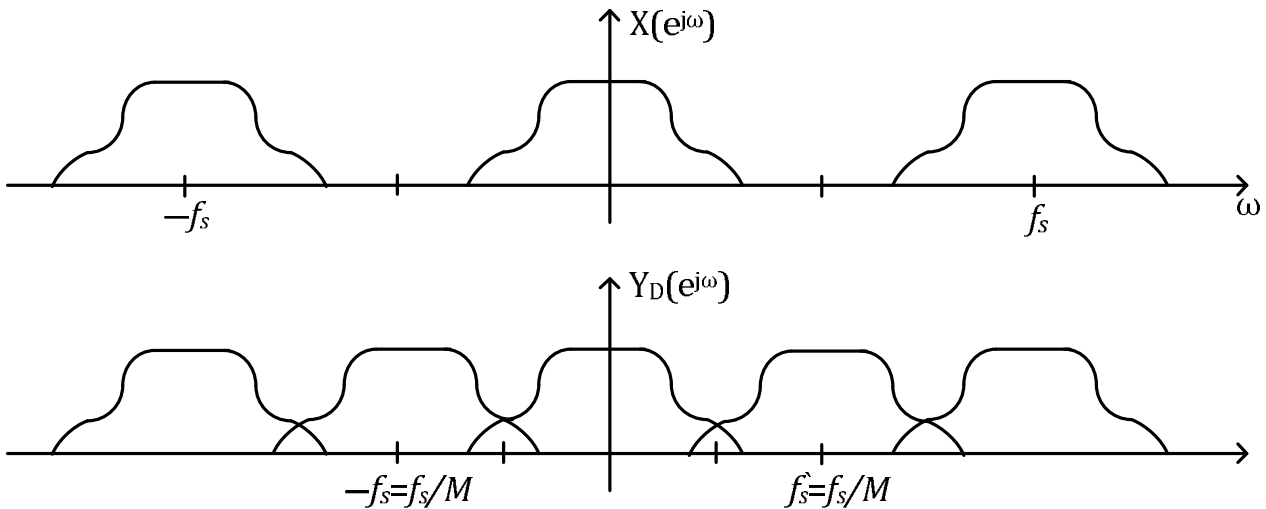


Рис. 2.3 – Изменение спектра сигнала после децимации ($M = 2$)

2.2.2 Интерполяция

Интерполятор используется в многоскоростной обработке сигналов для увеличения частоты дискретизации в L раз. Графическое изображение блока выполняющего интерполяцию показано на рис. 2.4. Интерполятор имеет следующее математическое описание

$$y_I(n) = \begin{cases} x(n/L), & n = 0, \pm L, \pm 2L \dots \\ 0 & \text{иначе} \end{cases}$$

либо можно записать эквивалентное выражение [2]

$$y_I(n) = \sum_{k=-\infty}^{\infty} x(k) \delta(n - kL),$$

из которого видно, что выходной сигнал получается путем подстановки $L - 1$ нуля между каждыми двумя последовательными отсчетами $x(n)$. Пример повышения частоты дискретизации показан на рис. 2.5.

В терминах z -преобразования интерполяция описывается уравнением

$$Y_I(z) = X(z^L), \quad (2.3)$$

из которого используя подстановку $z = e^{j\omega}$ получаем выражение для частотной характеристики сигнала после интерполяции:

$$Y_I(e^{j\omega}) = X(e^{j\omega L}). \quad (2.4)$$

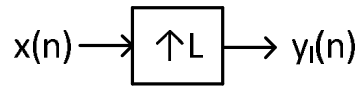


Рис. 2.4 – Блок-схема интерполятора

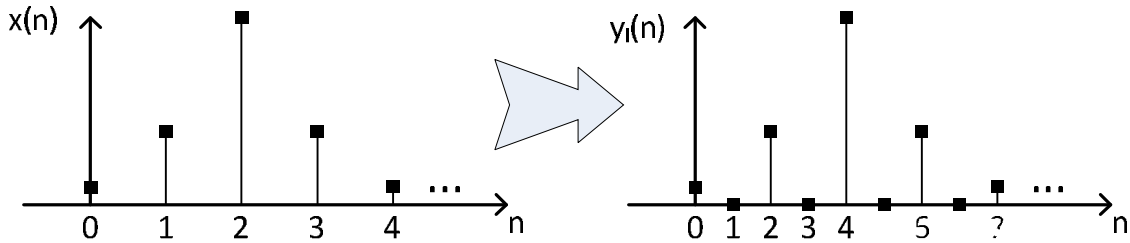


Рис. 2.5 – Блок-схема интерполятора ($L = 2$)

Последнее выражение свидетельствует о том, что спектр выходного сигнала есть в L раз сжатая версия спектра входного сигнала (происходит масштабирование частотной оси $\omega \mapsto \omega L$). На рис. 2.6 иллюстрируется изменение спектра сигнала после его интерполяции.

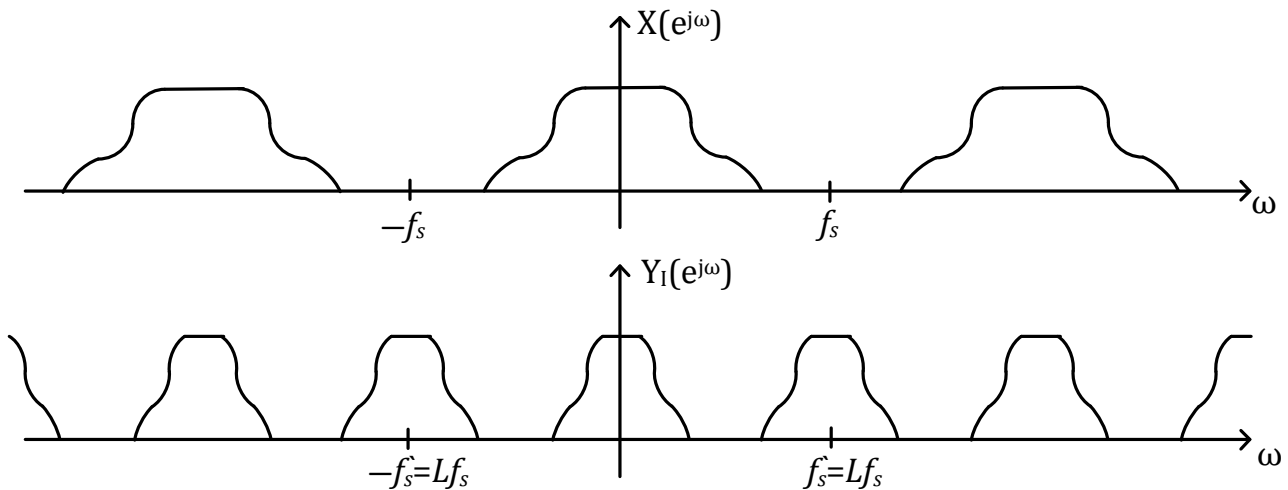


Рис. 2.6 – Интерполяция в частотной области ($L = 2$)

Из приведенного рисунка видно, что в процессе интерполяции в спектре сигнала появляются зеркальные частотные компоненты, которые отсутствовали в исходном сигнале. Для их удаления, как правило, используется на выходе интерполятора специальный фильтр нижних частот подавления зеркальных частот.

2.2.3 Эквивалентные схемы включения интерполяторов и дециматоров

В данном подразделе приводятся два «замечательных» тождества (от англ. – «Noble Identity»), которые оказываются весьма полезными для получения эффективных схем банков фильтров. Графическое изображение «замечательных» тождеств приведено на рис. 2.7.

Данные равенства (эквивалентные схемы включения) широко применяются для вывода различных соотношений в многоскоростной обработке сигналов. Эффективность тождественной схемы децимации обусловлена тем, что скорость поступления отсчетов на фильтр умень-

шается на коэффициент децимации, тем самым общая вычислительная сложность уменьшается пропорционально коэффициенту децимации (M). Аналогично и для тождественной схемы интерполяции фильтр работает на частоте в L раз меньшей, чем в исходной схеме.

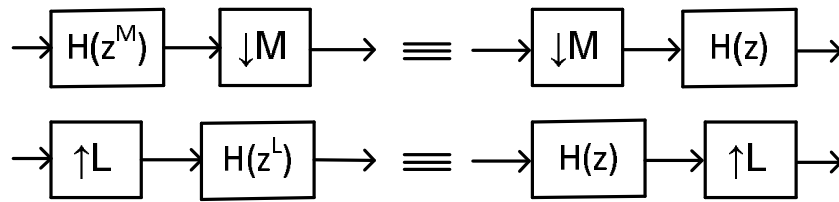


Рис. 2.7 – «Замечательные» тождества

2.3 Построение банков фильтров на основе модуляции

2.3.1 Основные понятия

Банком фильтров анализа называется совокупность фильтров $H_k(z)$, $0 \leq k < M$, которые раскладывают входной сигнал на M субполосных сигналов $x_k(n)$. Для случая $M = 4$ типичная частотная характеристика банка фильтров показана на рис. 2.8.

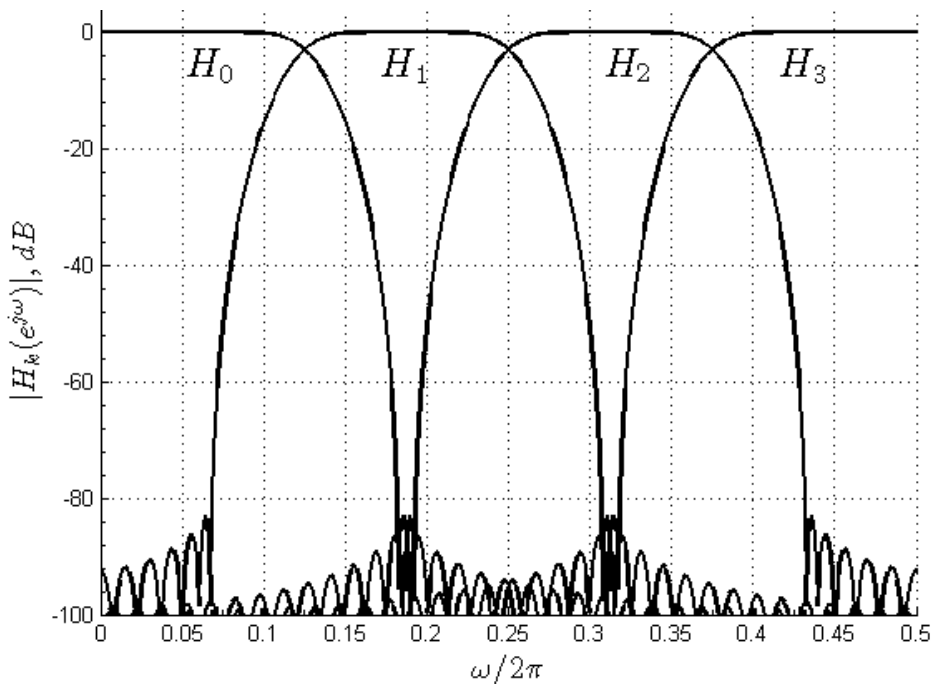


Рис. 2.8 – АЧХ 4-канального банка фильтров анализа

Банком фильтров синтеза называется набор фильтров $F_k(z)$, $0 \leq k < M$, которые объединяют M сигналов $v_k(n)$ в один сигнал $\hat{x}(n)$, который называют реконструированным. Соединение банка анализа и синтеза называют системой анализ/синтез, типовая структура которой изображена на рис. 2.9.

M -канальный банк фильтров, у которого все коэффициенты децимации/интерполяции одинаковы и равны M , называют предельно децимированным. Структура на рис. 2.9 позволяет вывести важное выражение для z -образа реконструированного (либо восстановленного) сигнала [3,4]:

$$\hat{X}(z) = \frac{1}{M} \sum_{k=0}^{M-1} \sum_{l=0}^{M-1} F_k(z) H_k(zW_M^l) X(zW_M^l). \quad (2.5)$$

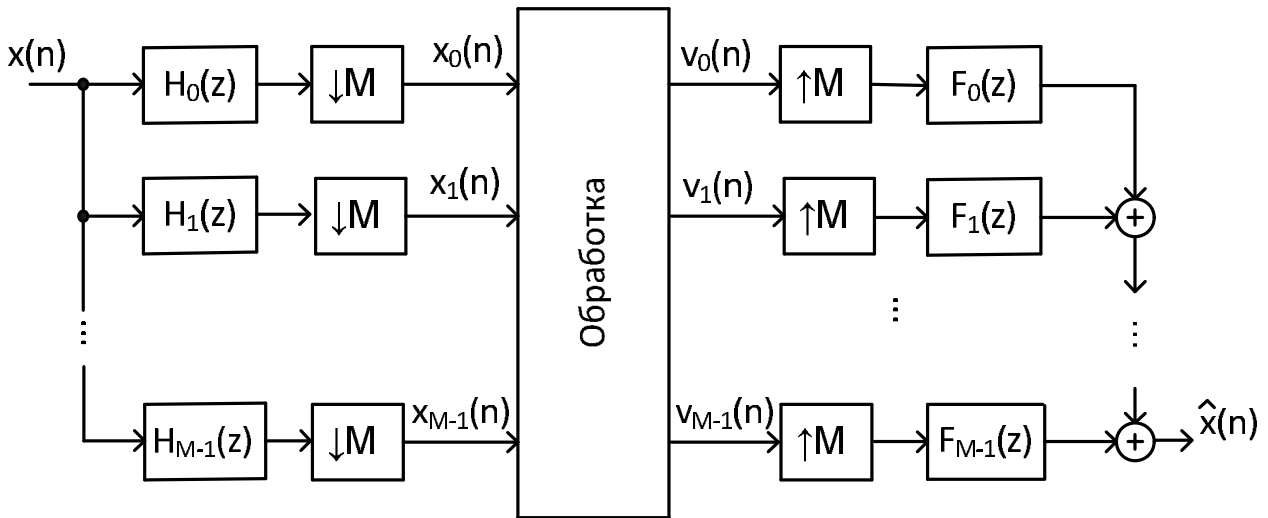


Рис. 2.9 – Структура системы анализ/синтез

Из данного выражения видно, что наряду с исходным сигналом восстановленный сигнал содержит $M - 1$ составляющих наложения. Появление данных составляющих есть негативное последствие децимации канальных сигналов (см. выражение (2.1)). Часто формулу (2.5) записывают в следующем виде [5]:

$$M\hat{X}(z) = X(z)T_0(z) + \sum_{l=1}^{M-1} X(zW_M^l)T_l(z), \quad (2.6)$$

где

$$T_l(z) = \sum_{k=0}^{M-1} F_k(z) H_k(zW_M^l), \quad 0 \leq l < M. \quad (2.7)$$

В выражении (2.6) $T_0(z)$ является передаточной функцией системы анализ/синтез, в то время как $T_l(z)$, $l \neq 0$ представляют собой передаточные функции компонентов наложения спектров, которые определяют, насколько хорошо подавляются компоненты наложения $X(zW_M^l)$.

Важным классом банков фильтров являются те, которые обладают свойством полного восстановления. Полное восстановление (от англ. «perfect reconstruction») означает, что сигнал на выходе схемы анализ/синтез (рис. 2.9) полностью идентичен сигналу на входе этой схемы с точностью до задержки, т.е. $\hat{x}(n) = x(n - D)$. Условие полного восстановления для системы анализ/синтез, состоящей из банка фильтров анализа $H_k(z)$ и банка фильтров синтеза $F_k(z)$, можно записать в матричной форме

$$\begin{bmatrix} H_0(z) & H_1(z) & \cdots & H_{M-1}(z) \\ H_0(zW_M) & H_1(zW_M) & \cdots & H_{M-1}(zW_M) \\ \vdots & \vdots & \ddots & \vdots \\ H_0(zW_M^{M-1}) & H_1(zW_M^{M-1}) & \cdots & H_{M-1}(zW_M^{M-1}) \end{bmatrix} \times \begin{bmatrix} F_0(z) \\ F_1(z) \\ \vdots \\ F_{M-1}(z) \end{bmatrix} = \begin{bmatrix} Mz^{-D} \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad (2.8)$$

Выражение (2.8) эквивалентно тому, как если бы мы задали следующие ограничения на передаточную функцию системы анализ/синтез и на передаточные функции компонент наложения спектров [6]:

$$\begin{cases} T_0(z) = z^{-D} \\ T_l(z) = 0, \end{cases} \quad 1 \leq l \leq M - 1. \quad (2.9)$$

Однако на практике проектирование банка фильтров, отвечающего таким жестким требованиям, как (2.8) является непростой задачей. Получаемые при этом банки фильтров, обладают не высоким уровнем затухания в полосе подавления (порядка 40–50 дБ см. [7]).

К счастью, во многих практических приложениях свойство полного восстановления не является критичным, поэтому условие (2.8) ослабляют, допуская небольшие колебания в передаточной функции системы анализ/синтез $T_0(z)$, получаемую таким образом систему называют банком фильтров с почти полным восстановлением (от англ. «near-perfect reconstruction»). Условие почти полного восстановления записывается как

$$\begin{cases} |T_0(e^{j\omega})| = 1 + \delta_1 \\ |T_l(e^{j\omega})| = \delta_2, \end{cases} \quad 1 \leq l \leq M - 1, \quad (2.10)$$

где δ_1 и δ_2 это величины малого порядка (≈ -100 дБ).

2.3.2 ДПФ-модулированный банк фильтров

2.3.2.1 Расчет фильтра-прототипа

Построение ДПФ-модулированного банка фильтров начинается с проектирования фильтра-прототипа $P(z)$, который является КИХ-фильтром нижних частот N -ого порядка с линейной ФЧХ и импульсной характеристикой $p(n)$

$$P(z) = \sum_{n=0}^{N-1} p(n)z^{-n}. \quad (2.11)$$

Частота среза фильтра-прототипа ω_c должна находится около частоты π/M , где M — это количество каналов банка фильтров. В некоторых случаях, когда хотят добиться лучшего подавления компонент наложения, для банков фильтров анализа и синтеза рассчитывают различные фильтры-прототипы. Однако в данном разделе рассматривается случай, когда в роли фильтра-прототипа для банков анализа и синтеза выступает один и тот же фильтр $P(z)$.

Импульсные характеристики фильтров анализа и синтеза записываются следующим образом:

$$h_k(n) = p(n)e^{j2\pi kn/M}, \quad (2.12)$$

$$f_k(n) = h_k^*(N - n - 1), \quad 0 \leq k < M, \quad (2.13)$$

где $h_k(n)$ — k -ый фильтр анализа, $f_k(n)$ — k -ый фильтр синтеза, N порядок фильтра-прототипа, и $*$ означает комплексное сопряжение.

Исходя из (2.12) можно записать z -преобразование для банка фильтров анализа:

$$H_k(z) = \sum_{n=0}^{N-1} h_k(n)z^{-n} = \sum_{n=0}^{N-1} p(n)(zW_M^{-k})^{-n} = P(zW_M^{-k}), \quad (2.14)$$

следовательно для частотной характеристики k -того фильтра банка анализа справедливо выражение

$$H_k(e^{j\omega}) = P\left(e^{j\left(\omega - \frac{2\pi k}{M}\right)}\right). \quad (2.15)$$

Выражение (2.15) говорит о том, что фильтры анализа задаются путем сдвига в частотной области АЧХ фильтра-прототипа. В результате мы получаем M -канальный банк фильтров, у которого амплитудно-частотные характеристики всех M фильтров анализа и синтеза это равномерно сдвинутые в частотной области версии фильтра-прототипа, поэтому подобный банк фильтров называют равнополосным.

Описываемый ниже метод расчета фильтра-прототипа для ДПФ-модулированного банка основывается на работе [5], где приводится метод расчета фильтра-прототипа для косинусно-модулированного банка фильтров. Таким образом, данный метод является расширением метода [5] на случай ДПФ-модулированного банка фильтров.

Для проектирования фильтра-прототипа вводится следующая функция ошибки:

$$E(\omega) = \begin{cases} \sum_{k=0}^{M-1} \left| P\left(\omega + \frac{2\pi k}{M}\right) \right|^2 - 1, & \omega \in \left[-\frac{\pi}{M}, \frac{\pi}{M}\right]. \\ P(\omega) & \omega \in [\omega_s, \pi] \end{cases} \quad (2.16)$$

В данном выражении ω_s частота отсечки полосы подавления фильтра-прототипа $P(\omega)$. Как правило, для задания частоты ω_s используется коэффициент перекрытия смежных полос [8] $0 < \rho < 1$ (от англ. «roll-of-factor»):

$$\omega_s = \frac{(1 + \rho)\pi}{M}. \quad (2.17)$$

Используя функцию ошибки формулируется взвешенная, среднеквадратичная целевая функция:

$$f = \sum_{\omega \in \left[-\frac{\pi}{M}, \frac{\pi}{M}\right] \cup [\omega_s, \pi]} B(\omega) |E(\omega)|^2, \quad (2.18)$$

где $B(\omega)$ функция стоимости (от англ. «cost function»), и $B(\omega) > 0 \forall \omega \in \left[-\frac{\pi}{M}, \frac{\pi}{M}\right] \cup [\omega_s, \pi]$.

Теперь задача расчета фильтра-прототипа сводится к минимизации функции (2.18). Оптимальное решение, которое минимизирует (2.18) трудно получить в аналитическом виде, поэтому для поиска минимума в [5] использовался алгоритм, который состоит из двух итеративных процедур, одна из которых находится внутри другой. Внутренний цикл итеративно оптимизирует коэффициенты фильтра-прототипа, в то время как во внешнем цикле процедуры итера-

тивно корректируется значения функции стоимости $B(\omega)$. В данном разделе рассматривается упрощенный вариант расчета, в котором $B(\omega) = 1$ и не изменяется на верхнем уровне алгоритма.

Индекс i обозначает значения функции или коэффициенты фильтра на i -той итерации. Частотная характеристика фильтра $P_i(\omega)$ записывается как

$$P_i(\omega) = e^{j(N-1)\omega} D_i(\omega), \quad (2.19)$$

где

$$D_i(\omega) = \mathbf{C}^T(\omega) \mathbf{d}_i, \quad (2.20)$$

и

$$\mathbf{d}_i = \left[h_i \left(\left\lfloor \frac{N}{2} \right\rfloor \right) \quad h_i \left(\left\lfloor \frac{N}{2} \right\rfloor + 1 \right) \quad \dots \quad h_i(N-1) \right]^T \quad (2.21)$$

$$\mathbf{C}(\omega) = \begin{cases} \left[2 \cos \left(\frac{\omega}{2} \right) \quad 2 \cos \left(\frac{3\omega}{2} \right) \quad \dots \quad 2 \cos \left(\frac{[N-1]\omega}{2} \right) \right]^T & \text{для четных } N \\ \left[1 \quad 2 \cos(\omega) \quad \dots \quad 2 \cos \left(\frac{[N-1]\omega}{2} \right) \right]^T & \text{для нечетных } N \end{cases}. \quad (2.22)$$

оператор $[a]$ обозначает наибольшее целое число меньше a .

Учитывая сделанные замечания, целевая функция принимает вид:

$$f_i = \sum_{\omega \in \left[-\frac{\pi}{M}, \frac{\pi}{M} \right] \cup [\omega_s, \pi]} |E_i(\omega)|^2, \quad (2.23)$$

Применяя выражение (2.20), функция ошибки (2.23) принимает следующий вид

$$E_i(\omega) = \begin{cases} \mathbf{d}_i^T \mathbf{P}(\omega) \mathbf{d}_i - 1 & \omega \in \left[-\frac{\pi}{M}, \frac{\pi}{M} \right] \\ \mathbf{C}^T(\omega) \mathbf{d}_i, & \omega \in [\omega_s, \pi], \end{cases} \quad (2.24)$$

где

$$\mathbf{Q}(\omega) = \mathbf{C}(\omega) \mathbf{C}^T(\omega),$$

$$\mathbf{P}(\omega) = \sum_{k=0}^{M-1} \mathbf{Q} \left(\omega + \frac{2k\pi}{M} \right).$$

Используя введенные обозначения, найдем первую производную целевой функции (2.23) по \mathbf{d}

$$\nabla_{\mathbf{d}} f_i = 4 \sum_{\omega \in \left[-\frac{\pi}{M}, \frac{\pi}{M} \right]} \{ \mathbf{d}_i^T \mathbf{P}(\omega) \mathbf{d}_i - 1 \} \mathbf{S}_i(\omega) + 2 \sum_{\omega \in [\omega_s, \pi]} \mathbf{R}_i(\omega),$$

где

$$\mathbf{R}_i(\omega) = D_i(\omega)\mathbf{C}(\omega),$$

$$\mathbf{S}_i(\omega) = \sum_{k=0}^{M-1} \mathbf{R}_i\left(\omega + \frac{2\pi k}{M}\right).$$

Пусть теперь $\mathbf{d}_{i,opt} = \mathbf{d}_i + \mathbf{e}_i$ является оптимальным решением, при котором $\nabla_{\mathbf{d}} f_i$ равно нулю. Чтобы определить \mathbf{e}_i разложим $\nabla_{\mathbf{d}} f_i$ в ряд Тейлора в окрестности точки оптимума:

$$\nabla_{\mathbf{d}} f_i \mathbf{d}_i = \underbrace{\nabla_{\mathbf{d}} f_i(\mathbf{d}_i + \mathbf{e}_i)}_0 - \nabla^2_{\mathbf{d}} f_i \mathbf{e}_i + \dots \quad (2.25)$$

где $\nabla^2_{\mathbf{d}} f_i$ - это матрица вторых производных (матрица Гессе) для функции f_i , которая имеет вид

$$\nabla^2_{\mathbf{d}} f_i = 4 \sum_{\omega \in \left[-\frac{\pi}{M}, \frac{\pi}{M}\right]} \{\mathbf{d}_i^T \mathbf{P}(\omega) \mathbf{d}_i - 1\} \mathbf{P}_i(\omega) + 8 \sum_{\omega \in \left[-\frac{\pi}{M}, \frac{\pi}{M}\right]} \mathbf{S}_i(\omega) \mathbf{S}_i^T(\omega) + 2 \sum_{\omega \in [\omega_s, \pi]} \mathbf{Q}(\omega). \quad (2.26)$$

Не учитывая члены высших порядков в (2.25), корректирующий вектор \mathbf{e}_i можно получить следующим образом:

$$\mathbf{e}_i = -(\nabla^2_{\mathbf{d}} f_i)^{-1} \nabla_{\mathbf{d}} f_i. \quad (2.27)$$

Таким образом, оптимизирующая процедура начинает работу с некоторого начального решения \mathbf{d}_0 , которое является грубым приближением оптимального. Далее на i -той итерации рассчитывается корректирующий вектор \mathbf{e}_i , по формуле (2.27), после чего вычисляются новые значения коэффициентов

$$\mathbf{d}_{i+1} = \mathbf{d}_i + \mathbf{e}_i.$$

Итерационная процедура заканчивается, когда $\|\mathbf{e}\|^2 \leq 10^{-10}$.

Для иллюстрации работы описанного метода, приведем пример расчета фильтра-прототипа для 16-канального ДПФ-модулированного банка фильтров. Порядок фильтра-прототипа равен 128, коэффициент перекрытия смежных полос $\rho = 0,9$. В качестве начального приближения выбираем КИХ-фильтр с частотой среза $\omega_c = \pi/M$, рассчитанный методом взвешивания на прямоугольное окно.

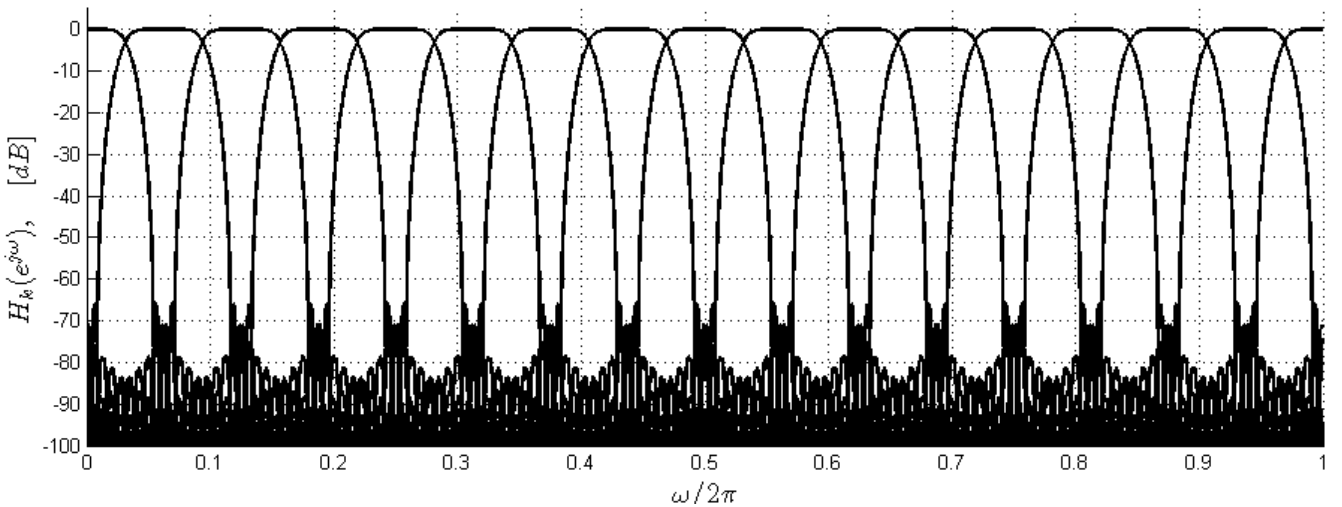


Рис. 2.10 – Частотная характеристика банка фильтров анализа

На рис. 2.10 показана частотная характеристика полученного банка фильтров анализа. Нужно отметить, что для вычисления коэффициентов фильтра-прототипа потребовалось двенадцать итераций оптимизационной процедуры. Ниже на рис. 2.11 приведена частотная характеристика передаточной функции анализ/синтез для полученного банка фильтров.

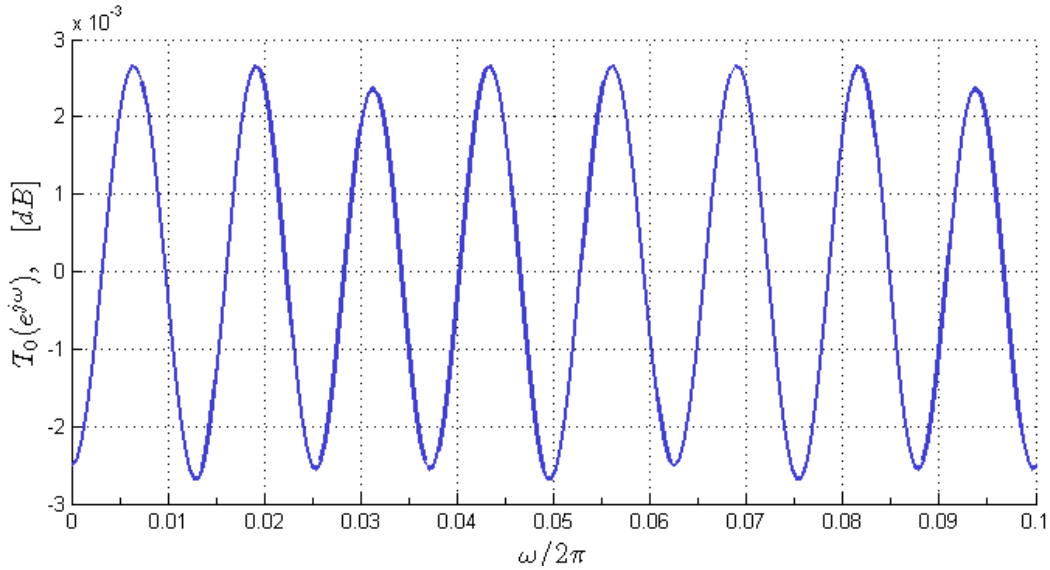


Рис. 2.11 – Передаточная функция $T_0(e^{j\omega})$ системы анализ/синтез

Из графика функции $T_0(e^{j\omega})$ можно сделать вывод, что банк фильтров не вносит в реконструированный сигнал искажений по уровню больших 0,003 дБ.

2.3.2.2 Полифазная структура ДПФ-модулированного банка фильтров

Первым шагом при получении полифазной структуры M -канального ДПФ-модулированного банка фильтров является полифазная декомпозиция фильтра-прототипа, которая заключается в разбиении его на M полифазных компонент:

$$P(z) = \sum_{l=0}^{M-1} z^{-l} E_l(z^M), 0 \leq l < M, \quad (2.28)$$

где

$$E_l(z) = \sum_{n=0}^{N/M-1} e_l(n) z^{-n} = \sum_{n=0}^{N/M-1} p(l + Mn) z^{-n}. \quad (2.29)$$

Другими словами импульсная характеристика $p(n)$ перегруппирована в M подпоследовательностей $e_l(n)$. Далее объединим выражения (2.14) и (2.29)

$$H_k(z) = \sum_{n=0}^{M-1} z^{-l} W_M^{-lk} E(z^M W_M^{kM}) = \sum_{l=0}^{M-1} z^{-l} W_M^{-kl} E_l(z^M). \quad (2.30)$$

В ходе вывода последнего выражения использовалось тождество $W_M^{kM} = 1$. Перепишем уравнение (2.30) в виде произведения вектора строки на вектор-столбец:

$$H_k(z) = \begin{bmatrix} 1 & W_M^{-k} & W_M^{-2k} & \dots & W_M^{-(M-1)k} \end{bmatrix} \begin{bmatrix} E_0(z^M) \\ z^{-1}E_1(z^M) \\ z^{-2}E_2(z^M) \\ \vdots \\ z^{-(M-1)}E_{M-1}(z^M) \end{bmatrix}. \quad (2.31)$$

Теперь все M уравнений (для каждого H_k) можно переписать в матричной форме:

$$\begin{bmatrix} H_0(z) \\ H_1(z) \\ H_2(z) \\ \vdots \\ H_{M-1}(z) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & W_M^{-1} & W_M^{-2} & \dots & W_M^{-(M-1)} \\ 1 & W_M^{-2} & W_M^{-4} & \dots & W_M^{-2(M-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_M^{-(M-1)} & W_M^{-2(M-1)} & \dots & W_M^{-(M-1)^2} \end{bmatrix} \begin{bmatrix} E_0(z^M) \\ z^{-1}E_1(z^M) \\ z^{-2}E_2(z^M) \\ \vdots \\ z^{-(M-1)}E_{M-1}(z^M) \end{bmatrix}. \quad (2.32)$$

Выражение (2.32) является ключевым для построения полифазной структуры ДПФ-модулированного банка фильтров, поскольку матрица $M \times M$ в правой части выражения есть не что иное, как матрица обратного дискретного преобразования Фурье (ОДПФ), умноженная на константу M . Это уравнение показывает путь к эффективной реализации банка фильтров анализа, структура которого приведена на рис. 2.12. Кстати, именно присутствие в чистом виде ДПФ в структуре банка фильтров и дало название данному классу банков фильтров.

Вычислительная сложность структуры на рис. 2.12 значительно меньше, чем у прямой реализации банка фильтров. Так для аппаратной реализации M -канального ДПФ-модулированного банка фильтров с фильтром-прототипом N -ого порядка понадобится $(M/2) \log_2 M + N$ операций умножения, в то время как для прямой реализации их потребуется $M \times N$.

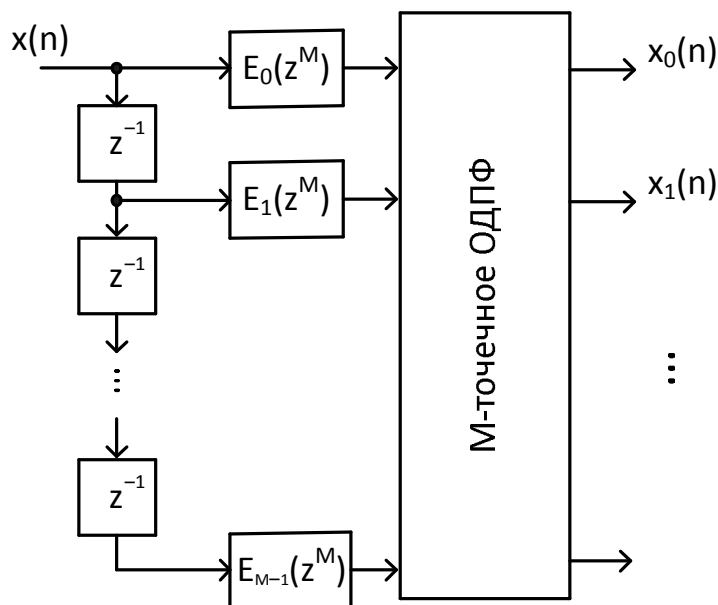


Рис. 2.12 – Структура ДПФ-модулированного банка фильтров анализа

При расчете числа операций умножения был учтен тот факт, что M -точечное ОДПФ эффективно вычисляется с помощью алгоритма Быстрого преобразования Фурье (БПФ).

Далее рассматривается построение полифазного банка фильтров синтеза. Исходной для нас будет являться структура, показанная на рис. 2.13. Фильтры $F_0(z), F_1(z) \dots F_{M-1}(z)$ должны быть схожими с $H_0(z), H_1(z) \dots H_{M-1}(z)$, только иметь другую фазу модуляции, которая обычно выбирается так, чтобы в результате импульсные характеристики фильтров синтеза была реверсными по времени (от англ. – «time-reverse») версиями фильтров анализа. В этом случае передаточная функция системы анализ/синтез будет иметь линейную ФЧХ. Для того чтобы фильтры синтеза удовлетворяли свойству реверсии по времени должно выполняться условие

$$F_k(z) = W_M^{-k} P(zW_M^k).$$

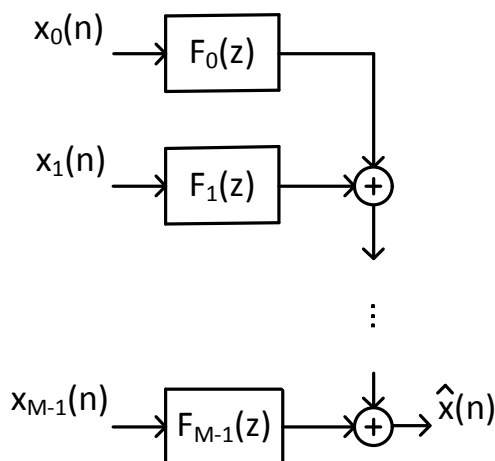


Рис. 2.13 – Банк фильтров синтеза

Вновь обратимся к полифазному представлению фильтра-прототипа:

$$P(z) = \sum_{l=0}^{M-1} z^{-(M-1-l)} R_l(z^M), \quad 0 \leq l < M, \quad (2.33)$$

где $R_l(z) = E_{M-1-l}(z)$. Комбинируя последние два выражения, получаем

$$F_k(z) = W_M^{-k} \sum_{l=0}^{M-1} z^{-(M-1-l)} W_M^{-(M-1-l)} R_l(z^M) = \sum_{l=0}^{M-1} z^{-(M-1-l)} W_M^{kl} R_l(z^M). \quad (2.34)$$

Далее выражение для банка фильтров синтеза можно представить в матричной форме:

$$\hat{X}(z) = [F_0(z) \quad F_1(z) \quad F_2(z) \quad \dots \quad F_{M-1}(z)] \begin{bmatrix} X_0(z) \\ X_1(z) \\ X_2(z) \\ \vdots \\ X_{M-1}(z) \end{bmatrix}. \quad (2.35)$$

Это же выражение можно записать в следующем виде:

$$\hat{X}(z) = [z^{-(M-1)}R_{M-1}(z^M) \quad z^{-(M-2)}R_{M-2}(z^M) \quad z^{-(M-3)}R_{M-3}(z^M) \quad \dots \quad R_0(z^M)] \times$$

$$\times \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & W_M^1 & W_M^2 & \dots & W_M^{(M-1)} \\ 1 & W_M^2 & W_M^4 & \dots & W_M^{2(M-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_M^{(M-1)} & W_M^{2(M-1)} & \dots & W_M^{(M-1)^2} \end{bmatrix} \times \begin{bmatrix} X_0(z) \\ X_1(z) \\ X_2(z) \\ \vdots \\ X_{M-1}(z) \end{bmatrix}. \quad (2.36)$$

Как и следовало ожидать, в выражении для банка синтеза присутствует матрица размера $M \times M$ прямого дискретного преобразования Фурье. Таким образом, мы получили выражение, в соответствии с которым строится эффективная полифазная структура банка фильтров синтеза (рис. 2.14).

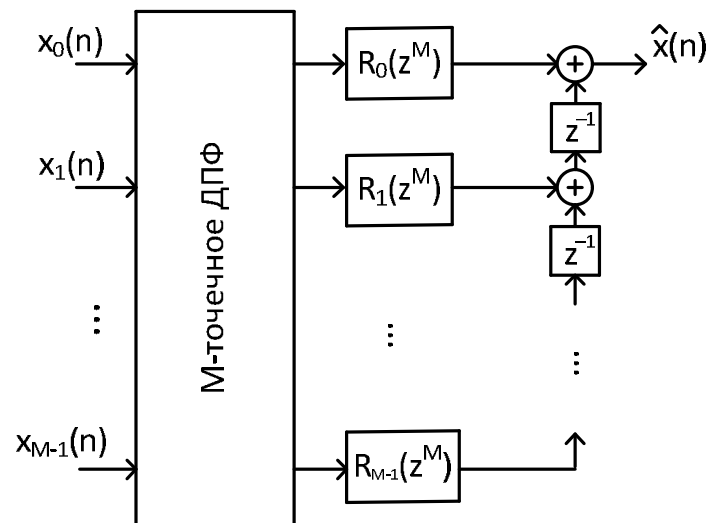


Рис. 2.14 – Структура ДПФ-модулированного банка фильтров синтеза

Вычислительная сложность банка фильтров синтеза такая же, как у банка фильтров анализа. Ниже приведена реализация полифазного предельно децимированного ДПФ-модулированного банка фильтров в среде MATLAB.

```
function [X, XF, XK, hp] = dft02a(x, K, h)
%Максимально децимированный ДПФ-модулированный банк фильтров анализа
%Вход:
% x - анализируемый сигнал
% K - количество каналов
% h - коэффициенты фильтра-прототипа (ИХ) Low Pass FIR (wc =
% pi / K)
%Выход:
% X - матрица с сигналами каналов банка фильтра после ОДПФ
% XF - сигналы каналов после полифазной фильтрации
% XK - сигналы каналов после децимации
% hp - матрица полифазных фильтров

%Сортировка входных отсчетов сигнала на каналы - децимация
X = x;
if rem(length(X), K),
    X(K * ceil(length(X) / K)) = 0;
```

```

end
X = reshape(X, K, length(X) / K);
if (nargout > 2),           XK = X;           end

%Разбиение фильтра-прототипа на полифазные фильтры
if mod(length(h), K),
    h(K * ceil(length(h) / K)) = 0;
end
h = flipud(reshape(h, K, length(h) / K));
if (nargout > 3),           hp = h;           end

%Полифазная фильтрация каналов
for i = 1 : K,
    X(i, :) = filter(h(i, :), 1, X(i, :));
end
if (nargout > 2),           XF = X;           end

%Модуляция посредством ОДФ
X = K*iFFT(X);

```

function [x, xF, xIDFT, hp] = dft02s(X, h)

%Максимально децимированный ДПФ-модулированный банк фильтров синтеза

%Вход:

% X - матрица с сигналами каналов
% h - коэффициенты фильтра-прототипа (ИХ) Low Pass FIR (wc =
% pi / K)

%Выход:

% x - синтезированный сигнал
% xF - сигналы в каналах после фильтрации
% xDFT - сигналы в каналах после ДПФ
% hp - матрица полифазных фильтров

%Определение количества каналов

K = size(X, 1);

%ДПФ

x = fft(X);

if (nargout > 2), xIDFT = x; end

%Разбиение фильтра-прототипа на полифазные фильтры

if rem(length(h), K),
 h(K * ceil(length(h) / K)) = 0;

end

h = reshape(h, K, length(h) / K) * K * K;

if (nargout > 3), hp = h; end

%Полифазная фильтрация каналов

for i = 1 : K,
 x(i, :) = filter(h(i, :), 1, x(i, :));

end

if (nargout > 1), xF = x; end

%Синтезированный сигнал

x = x(:);

2.3.3 Неравнополосный ДПФ-модулированный банк фильтров

В данном разделе будет рассмотрено построение неравнополосного ДПФ-модулированного банка фильтров на основе фазового преобразования (то англ. «allpass transform»). Выбирается фазовое звено первого порядка, которое имеет следующую передаточную функцию

$$A(z) = \frac{z^{-1} + \alpha}{1 + \alpha z^{-1}} \quad (2.37)$$

Фазовое звено не изменяет амплитуды частотных компонент, его АЧХ равно единице для всех частот. Однако ФЧХ фазового звена нелинейна и зависит от параметра α [9]

$$\varphi(\omega) = -\omega - 2\text{arctg} \frac{\alpha \sin \omega}{\alpha \cos \omega - 1} \quad (2.38)$$

При $\alpha = 0$ фазовое звено превращается в простой элемент задержки z^{-1} . Нелинейность ФЧХ вызывает групповая задержки, что приводит к тому, что компоненты сигнала в зависимости от частоты проходят через фазовое звено медленнее или быстрее. Как показано в [10], если сигнал пропустить через цепь фазовых звеньев (рис. 2.15), что его спектр будет претерпевать деформации.

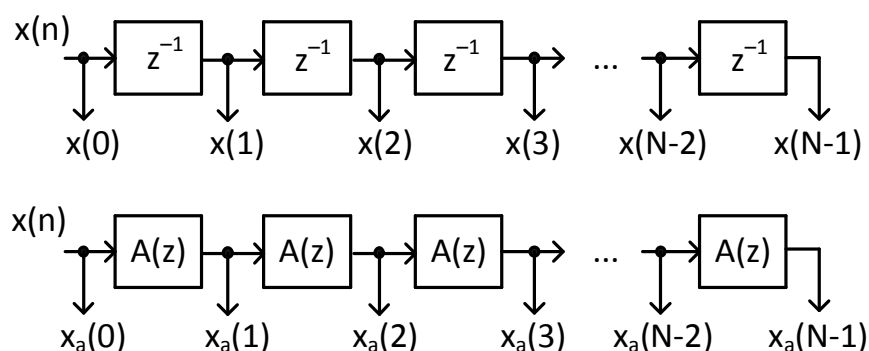


Рис. 2.15 – Механизм фазового преобразования

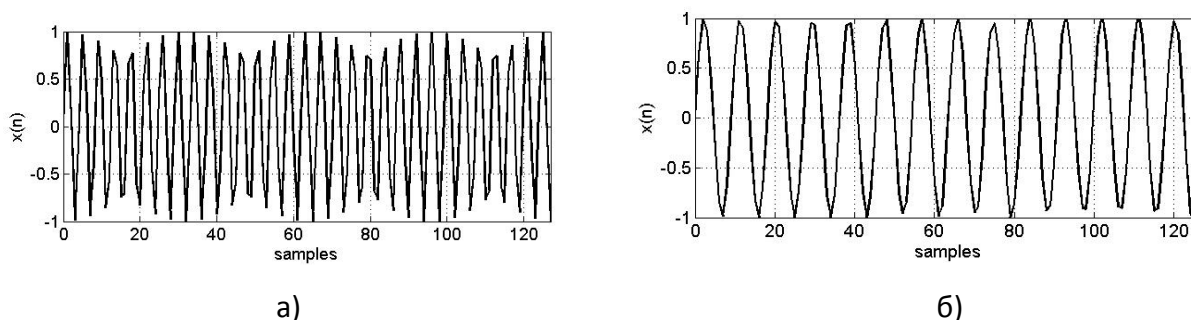


Рис. 2.16 – Прохождение сигнала через цепь фазовых звеньев

На рис. 2.16 приведен пример прохождения синусоиды через первую и вторую цепь на рис. 2.15 (параметр $\alpha = -0.45$). Интереснее посмотреть, как изменился спектр этого сигнала. На рис. 2.17 показано, как нелинейная ФЧХ повлияла на изменение спектра сигнала.

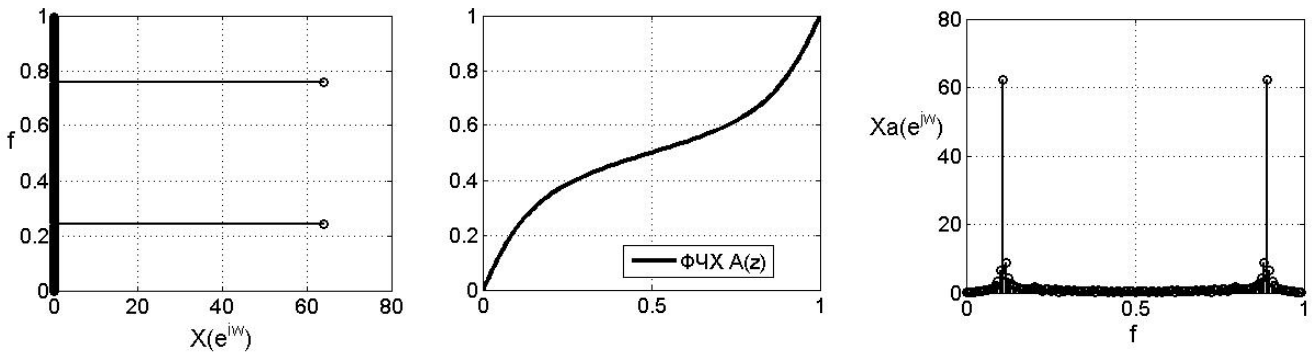


Рис. 2.17 – Суть фазового преобразования

Таким образом, замена всех элементов задержки на фазовые звенья привела к тому, что в частотной области происходит трансформация частотной оси $\omega \rightarrow \varphi(\omega)$, или другими словами происходит отображение оси ω на ось $\varphi(\omega)$, как показано на рис. 2.17.

2.3.4 Полифазная структура неравнополосного ДПФ-модулированного банка фильтров

В данном разделе будет рассмотрена полифазная структура ДПФ-модулированного банка фильтров, которая впервые была описана в [9]. Неравнополосная версия ДПФ-модулированного банка фильтров получается из равнополосной путем замены всех элементов задержки на фазовые звенья $z^{-1} \rightarrow A(z)$.

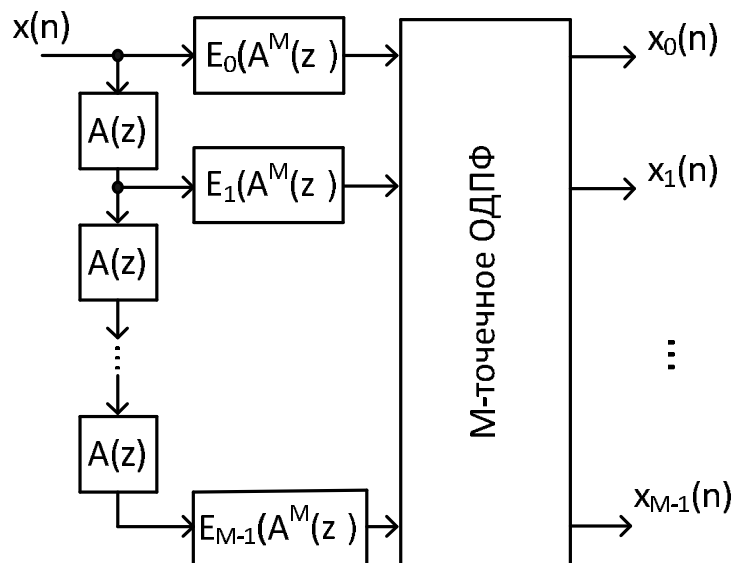


Рис. 2.18 – Полифазная структура неравнополосного ДПФ-модулированного банка фильтров анализа

В схеме на рис. 2.18 $E_r(A^M(z)), r = 0, \dots, M - 1$ это, подвергнутый фазовой трансформации, полифазный компонент, который определяется как

$$E_r(A^M(z)) = \sum_{l=0}^{M-1} p(lM + r)A^{lM}(z), \quad (2.39)$$

где $p(n)$ это коэффициенты фильтра-прототипа, который может быть рассчитан теми же методами, что и для равнополосного банка фильтров. Согласно рис. 2.18 z -преобразование для k -того фильтра анализа неравнополосного ДПФ-модулированного банка фильтров имеет вид:

$$H_k(z) = \sum_{r=0}^{M-1} E_r(A^M(z))A^r(z)W_M^{-kr}. \quad (2.40)$$

Аналогичным образом строится неравнополосный ДПФ-модулированный банк фильтров синтеза (рис. 2.19).

Фильтр $C(z)$ необходим для того, чтобы сделать ФЧХ системы анализ/синтез линейной, при этом выравнивается групповая задержка, что является важным, если предполагается использовать банк фильтров для обработки речевых сигналов. Методика проектирования корректирующего фильтра $C(z)$ рассмотрена в [11].

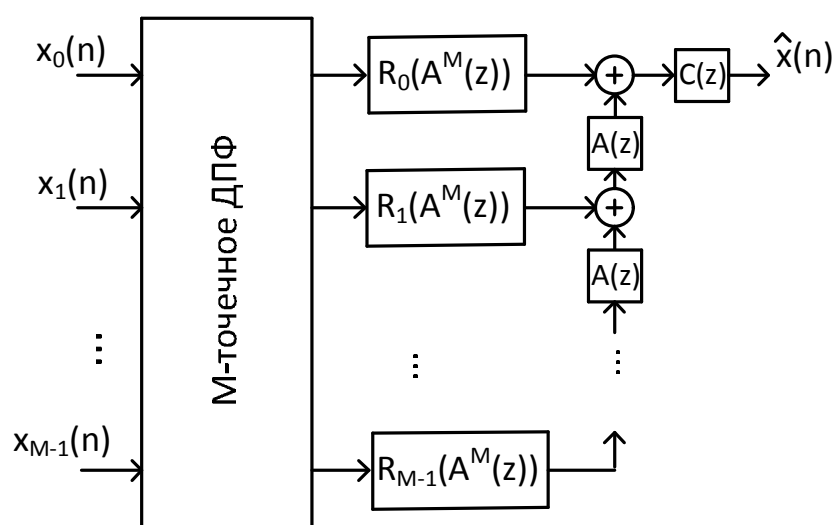


Рис. 2.19 – Полифазная структура неравнополосного ДПФ-модулированного банка фильтров синтеза

Ниже приводятся MATLAB-реализации неравнополосного ДПФ-модулированного банка анализа и синтеза.

```
function [X] = dft04a(x, K, M, h, a, sm, sc)
%Обобщенный ДПФ-модулированный неравнополосный банк фильтров анализа
%Вход:
% x - анализируемый сигнал
% K - количество каналов
% M - коэффициент децимации канальных сигналов
% h - коэффициенты фильтра-прототипа Low Pass FIR (wc = pi/K)
% a - коэффициенты фазового фильтра
% sm - признак однократного перекрытия «1» или «0»
% sc - компенсация перекрытия полос «1» или «0»
%Выход:
% X - матрица с сигналами каналов банка фильтра
if (nargin < 7),
    sc = 1;
end

%Расчет коэффициентов всепропускающего фильтра
```

```

ap_a = 1;
ap_b = 1;
ap_K = length(a);
for i = 1 : ap_K,
    ap_a = conv(ap_a, [1 -a(i)]);
    ap_b = conv(ap_b, [-conj(a(i)) 1]);
    if sc,
        ap_abs = abs(a(i));
        ap_ang = angle(a(i));
        ap_fi = -2*atan(ap_abs*sin(ap_ang) / (1-ap_abs*cos(ap_ang)));
        ap_b = ap_b * exp(j * ap_fi);
    end;
end;
end;

```

%Подготовка фильтра-прототипа

```
lp_L = length(h);
```

```
lp_h = h * M * K;
```

%Фильтрация сигнала

```

if sm,
    X = zeros(K, ceil((length(x) + (lp_L - 1) * (ap_K - 1)) / M));
    tmp = x;
    for i = 1 : lp_L,
        qq = (lp_L - 1) * (ap_K - 1);
        pp = (lp_L - i) * (ap_K - 1);
        tmp2 = [zeros(1, pp) tmp zeros(1, qq - pp)];
        tmp2 = tmp2(1 : M : length(tmp2)) * lp_h(i);
        X(mod(i - 1, K) + 1, :) = X(mod(i - 1, K) + 1, :) + tmp2;
        tmp = filter(ap_b, ap_a, tmp);
    end
else
    X = zeros(K, ceil(length(x) / M));
    tmp = x;
    for i = 1 : lp_L,
        tmp2 = tmp(1 : M : length(tmp)) * lp_h(i);
        X(mod(i - 1, K) + 1, :) = X(mod(i - 1, K) + 1, :) + tmp2;
        tmp = filter(ap_b, ap_a, tmp);
    end
end
end

```

%Модуляция при помощи ОДПФ

```
X = ifft(X);
```

function [x] = dft04s(X, M, h, a, sm, sc)

%Обобщенный ДПФ-модулированный неравнополосный банк фильтров синтеза

%Вход:

```

% X      -   матрица с сигналами каналов
% K      -   количество каналов
% M      -   коэффициент децимации каналов банка фильтра
% h      -   коэффициенты фильтра-прототипа Low Pass FIR (wc = pi/K)
% a      -   коэффициенты фазового фильтра
% sm     -   признак однократного перекрытия «1» или «0»
% sc     -   компенсация перекрытия полос «1» или «0»

```

```

%Выход:
%   x   -   синтезированный сигнал
if (nargin < 6),          sc = 1;          end

%Определение количества каналов
K = size(X, 1);

%Модуляция при помощи ДПФ
X = fft(X);

%Расчет коэффициентов всепропускающего фильтра
ap_a = 1;
ap_b = 1;
ap_K = length(a);
for i = 1 : ap_K,
    ap_a = conv(ap_a, [1 -a(i)]);
    ap_b = conv(ap_b, [-conj(a(i)) 1]);
    if sc,
        ap_abs = abs(a(i));
        ap_ang = angle(a(i));
        ap_fi = -2 * atan(ap_abs * sin(ap_ang) / ...
            (1 - ap_abs * cos(ap_ang)));
        ap_b = ap_b * exp(j * ap_fi);
    end;
end;

%Подготовка фильтра-прототипа
lp_L = length(h);
lp_h = 1 * h;

%Фильтрация сигнала
if sm,
    tmp1(1 : M : size(X, 2) * M) = X(1, :) * lp_h(1);
    x = [tmp1 zeros(1, (lp_L - 1) * (ap_K - 1))];
    for i = 2 : lp_L,
        x = filter(ap_b, ap_a, x);
        qq = (lp_L - 1) * (ap_K - 1);
        pp = (lp_L - i) * (ap_K - 1);
        tmp1(1:M:size(X, 2)*M) = X(mod(i - 1, K) + 1, :) * lp_h(i);
        x = x + [zeros(1, qq - pp) tmp1 zeros(1, pp)];
    end;
else
    x = zeros(1, length(X) * M);
    x(1 : M : length(x)) = X(1, :) * lp_h(1);
    for i = 2 : lp_L,
        x = filter(ap_b, ap_a, x);
        x(1 : M : length(x)) = x(1 : M : length(x)) + ...
            X(mod(i - 1, K) + 1, :) * lp_h(i);
    end;
end;
end;

```

2.3.5 Примеры проектирования неравнополосных ДПФ-модулированных банков фильтров

В данном подразделе проиллюстрированы некоторые практические аспекты расчета неравнополосных ДПФ-модулированных банков фильтров. В начале, используя методику, изложенную в параграфе 2.3.2.1, рассчитывается 32-канальный равнополосный ДПФ-модулированный банк фильтров. При этом задается порядок фильтра-прототипа $N = 192$, и коэффициент перекрытия смежных полос $\rho = 0,75$. Полученный результат расчета банка фильтров показан на рис. 2.20.

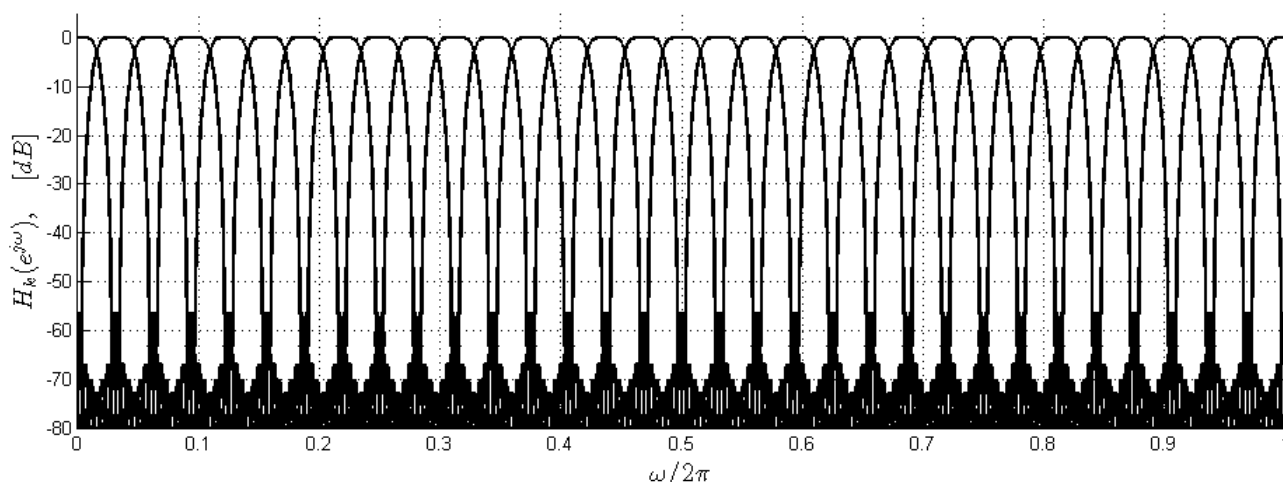


Рис. 2.20 – АЧХ равнополосного 32-канального ДПФ-модулированного банка фильтров

АЧХ системы анализ/синтез, которая отражает уровень вносимых искажений, показана на рис. 2.21. Из данного графика видно, что банк фильтров не вносит искажений превышающих по уровню 0,008 дБ.

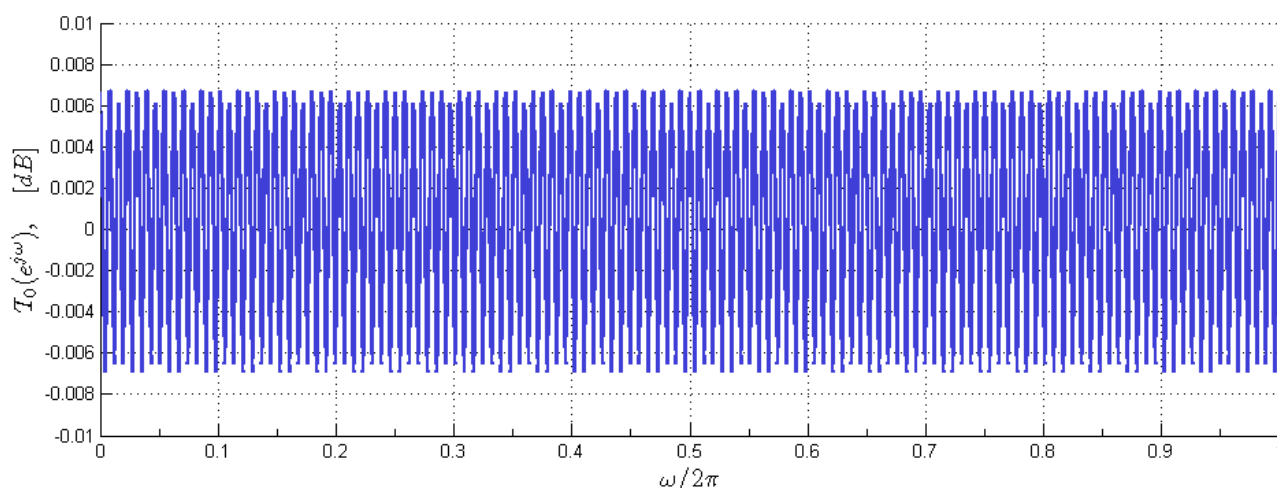


Рис. 2.21 – АЧХ системы анализ/синтез

Кроме того, система обладает линейной ФЧХ и постоянной групповой задержкой, равной $\tau = N - 1 = 191$ отсчету. Далее, к полученному банку фильтров применяется фазовое преобразование, которое, как упоминалось ранее, состоит в замене всех элементов задержки на фазовые звенья: $z^{-1} \rightarrow A(z) = (z^{-1} + \alpha)/(1 + \alpha z^{-1})$. Это преобразование приводит к деформации частотной оси и изменению ширины полос пропускания субфильтров в банке фильтров

(рис. 2.22). АЧХ системы анализ/синтез с деформированной частотной осью показана на рис. 2.23. Негативный эффект от использования фазового преобразования можно увидеть на графике ФЧХ системы анализ/синтез с деформированной частотной осью (рис. 2.24). ФЧХ системы становится нелинейной, что весьма нежелательно при обработке речевых и звуковых сигналов.

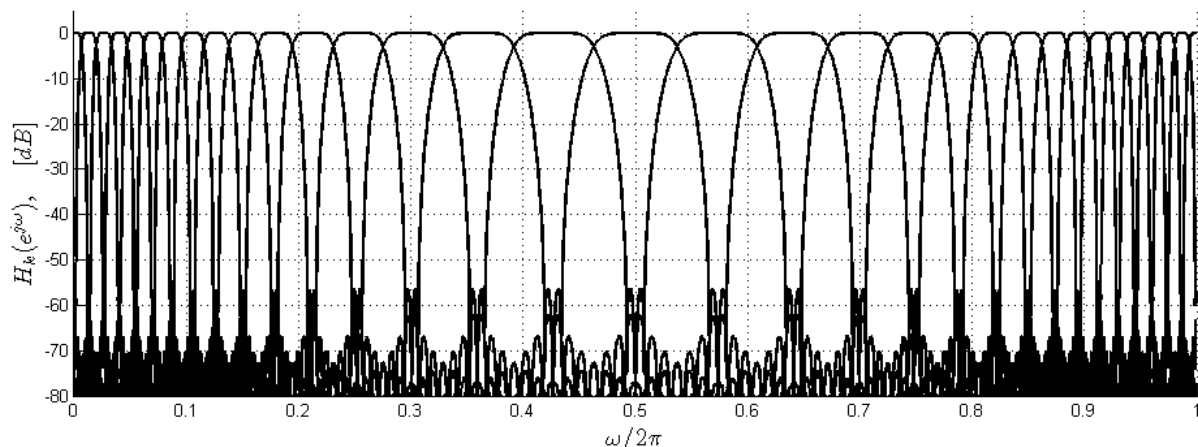


Рис. 2.22 – АЧХ неравнополосного 32-канального ДПФ-модулированного банка фильтров анализа ($\alpha = -0.4092$)

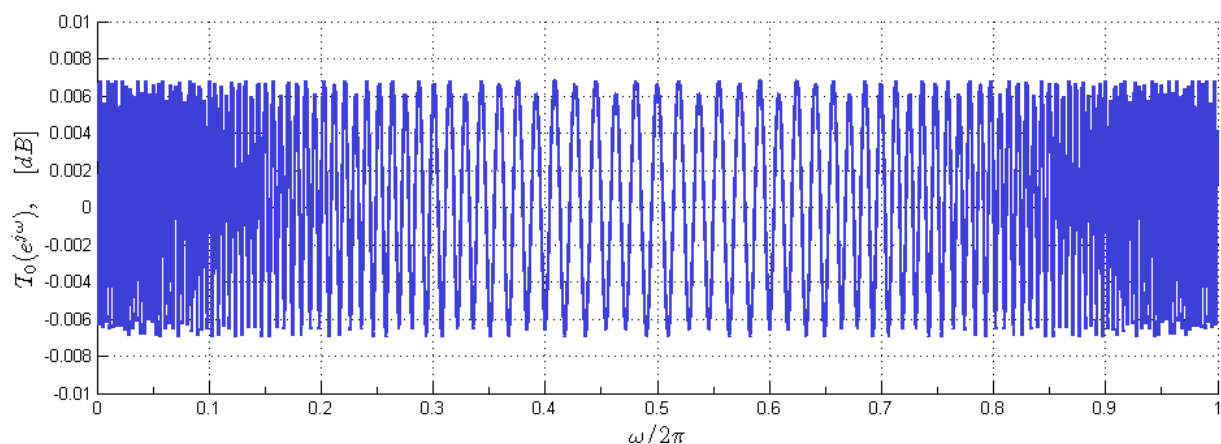


Рис. 2.23 – АЧХ системы анализ/синтез ($\alpha = -0.4092$)

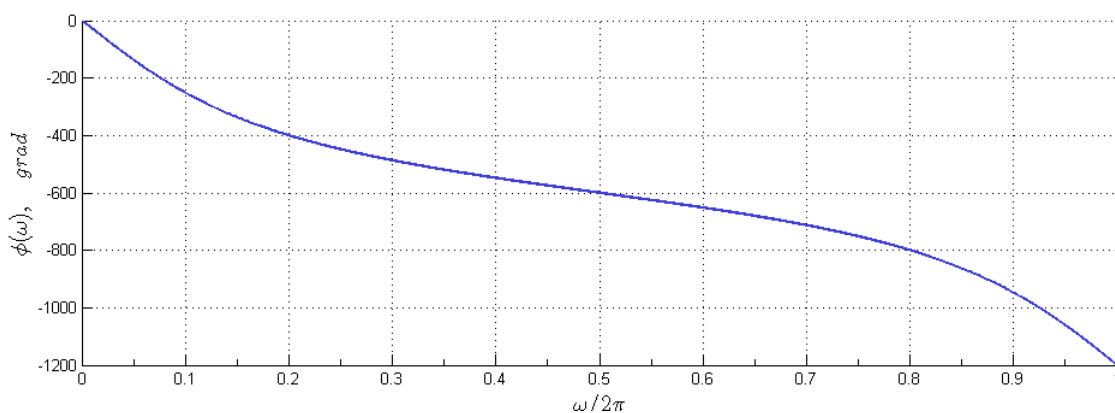


Рис. 2.24 – ФЧХ системы анализ/синтез ($\alpha = -0.4092$)

Как следствие нелинейности ФЧХ, групповая задержка системы анализ/синтез перестает быть постоянной функцией (рис. 2.25). Для того, чтобы скорректировать ФЧХ в работе [11]

был предложен метод проектирования специального пост-фильтра $C(z)$ с конечной импульсной характеристикой. Структурно данный фильтр располагается сразу после банка фильтров синтеза, как показано на рис. 2.19. Графики ФЧХ и групповой задержки системы анализ/синтез с использованием фильтра $C(z)$ порядка $N_{eq} = 422$ показаны на рис. 2.26 и рис. 2.27, соответственно.

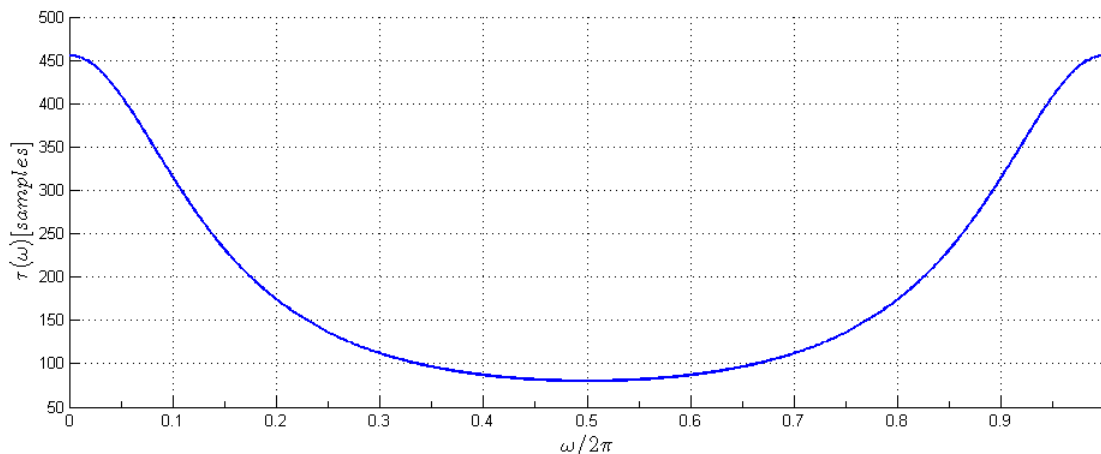


Рис. 2.25 – Групповая задержка системы анализ/синтез ($\alpha = -0.4092$)

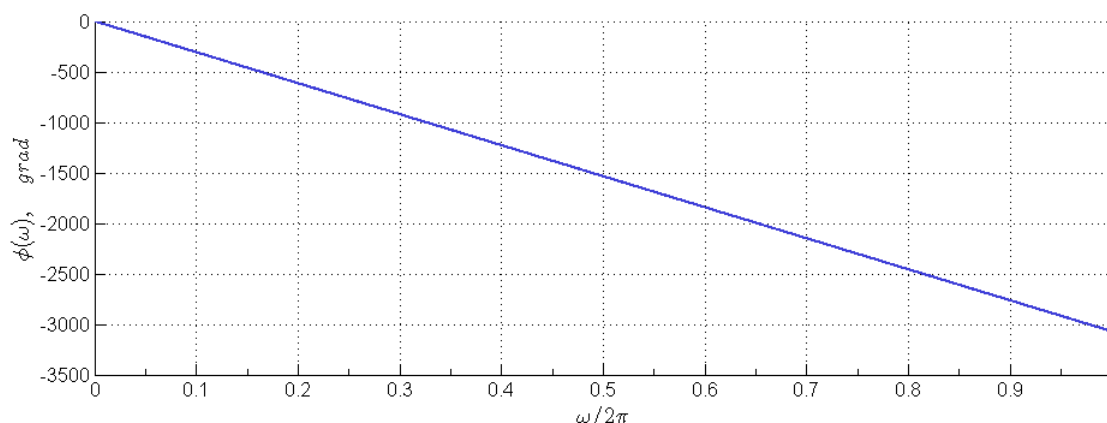


Рис. 2.26 – Скомпенсированная ФЧХ системы анализ/синтез ($\alpha = -0.4092$)

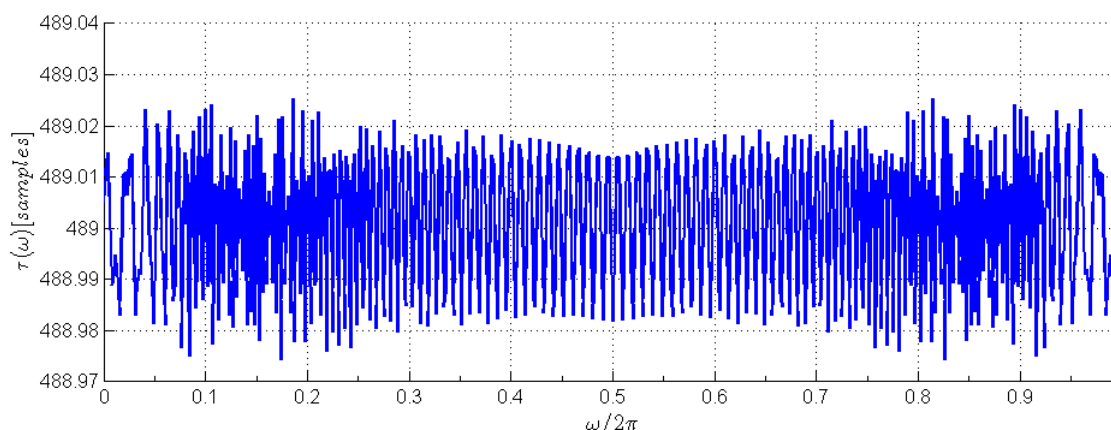


Рис. 2.27 – Скомпенсированная групповая задержка системы анализ/синтез ($\alpha = -0.4092$)

Таким образом, в разделе рассмотрены некоторые практические аспекты расчета неравнополосных ДПФ-модулированных банков фильтров. Следует отметить, что применение

банков фильтров основанных на фазовом преобразовании для обработки речевых и звуковых сигналов возможно лишь при условии использования специального фильтра выравнивающего групповую задержку системы анализ/синтез.

2.4 Список использованных источников

1. Айфичер, Э. Цифровая обработка сигналов: практический подход, 2-е издание; пер. с англ. — М. : Издательский дом «Вильямс», 2004. — 992 с.
2. Vaidyanathan, P.P. "Multirate digital filters, filter banks, polyphase networks and application: a tutorial", *Proceedings of the IEEE*, vol. 78, no. 1, pp. 5693, 1990.
3. Витязев, В.В. "Многоскоростная обработка сигналов: ретроспектива и современное состояние (часть 1)", *Цифровая обработка сигналов, №1*, с. 12–21, 2008.
4. Piotrowski A., Parfieniuk M., "Digital filter banks: analysis, synthesis and implementation for multimedia systems" *Wydawnictwo Politechniki Bialostockiej, Bialystok*, 2006.
5. Goh C.K., "An efficient algorithm for the design of weighted minimax M -channel cosine-modulated filter bank", *IEEE Trans. On signal processing*, vol. 46, no. 5, pp. 1426 – 1430, 1998.
6. Nguyen, T. Q. "Near-perfect-reconstruction pseudo-QMF banks", *IEEE Transactions on signal processing*, vol. 42, No. 1, pp. 65 – 76, 1998.
7. Koilpillai, D., Vaidyanathan P.P. "Cosine-Modulated FIR Filter Banks Satisfying Perfect Reconstruction", *IEEE Trans. on Signal Processing*, vol. 40, no. 40. – pp. 770-783, 1992.
8. Furtado M.B., Diniz P.S., Netto S.L. "Numerically efficient optimal design of cosine-modulated filter banks with peak-constrained least-squares behavior", *IEEE transactions on circuits and systems*, vol. 52, no. 3, – pp. 597 – 608, 2006.
9. Galijasevic E., J. Kliewer "Non-uniform near-perfect-reconstruction oversampled DFT filter banks based on all-pass transform", *The Proc. of IEEE Digital Signal Processing Workshop*, 2000 pp. 1–6.
10. Oppenheim A., Johnson D., Streiglitz K. "Computation of Spectra with Unequal Resolution Using the Fast Fourier Transform", *Proceeding of IEEE*, vol. 59, no 2. – pp. 299-301, 1971.
11. Parfieniuk M., Petrovsky A., Wan W. "Frequency warping and subband merging for approximating the critical bands with cosine modulated filter banks", *Internationals conference on audio, language and image processing (ICALIP)*, 2008, vol.2, pp. 1159–1166.

3 ПАРАУНИТАРНЫЕ БАНКИ ФИЛЬТРОВ НА ОСНОВЕ АЛГЕБРЫ КВАТЕРНИОНОВ

Парфенюк М., Петровский А.А.

АННОТАЦИЯ

Параунитарные банки фильтров (ПУБФ) могут рассматриваться как наиболее значимое преобразование среди многоскоростных систем цифровой обработки сигналов [1-5]. Это обусловлено тем фактом, что подобные банки являются преобразованиями без потерь в дополнение к гарантированной перфективной реконструкции сигнала. Точное соотношение между энергиями во всем частотном диапазоне и в субполасах сильно упрощает теоретические выкладки и делает ПУБФ полезными в применении к компрессии (кодированию) изображений. Серьезной практической проблемой, связанной с реализацией ПУБФ на арифметике с фиксированной запятой, является потеря свойства перфективной реконструкции сигнала банком фильтров. Это обусловлено чувствительностью структуры ПУБФ к квантованию коэффициентов. В данной работе представлен новый подход к синтезу и конструированию решетчатых структур 4-х и 8-ми канальных ПУБФ на основе алгебры кватернионов. В отличие от известных решений, предложенные структуры являются структурами без потерь даже при выполнении вычислений с конечной точностью.

Финансовая поддержка работы осуществлялась Министерством науки и высшей школы Республики Польша (грант № 519 030 32/3775)

Ключевые слова – банки цифровых фильтров, параунитарные банки фильтров, перфективная реконструкция сигнала, алгебра кватернионов.

СОДЕРЖАНИЕ

3.1	Решетчатые структуры параунитарных банков фильтров.....	50
3.2	Регулярность и эффекты квантования коэффициентов.....	51
3.3	Кватернионы и ортогональные матрицы.....	52
3.3.1	Кватернионы.....	52
3.3.2	Матрицы умножения.....	53
3.3.3	Параметризация ортогональных матриц.....	54
3.3.4	Умножитель кватернионов как составной блок ПУБФ.....	55
3.4	Решетчатые структуры ПУБФ на основе кватернионов.....	56
3.4.1	Четырех канальный ПУБФ.....	56
3.4.2	Восьми канальный ПУБФ.....	60
3.5	Примеры синтеза.....	63
3.5.1	Процедура расчета.....	63
3.5.2	Четырех канальные ПУБФ.....	63
3.5.3	Восьми канальные ПУБФ.....	65
3.6	Заключение.....	71
3.7	Список использованных источников.....	71

3.1 Решетчатые структуры параунитарных банков фильтров

M -канальный параунитарный банк фильтров анализа в общем виде может быть описан следующей факторизацией полифазной передаточной матрицы [6]:

$$\mathbf{E}(z) = \mathbf{R}_N(z)\mathbf{\Lambda}(z)\mathbf{R}_{N-1}\mathbf{\Lambda}(z) \dots \mathbf{R}_1\mathbf{\Lambda}(z)\mathbf{E}_0, \quad (3.1)$$

которая содержит $\mathbf{\Lambda}(z) = \text{diag}(z^{-1}, \mathbf{I}_{M-1})$ - матрицу задержек, ортогональные матрицы: общую \mathbf{E}_0 с $M(M-1)/2$ степенями свободы и $\mathbf{R}_i, i = 1 \dots N-1$, ограниченные до $M-1$ степени свободы. Оба типа матриц параметризованы планарными оборотами Гивенса, в которых каждый из них соответствует одной степени свободы.

Банк фильтров анализа, определяемый выражением (3.1), может быть представлен следующей решетчатой структурой (рис. 3.1).

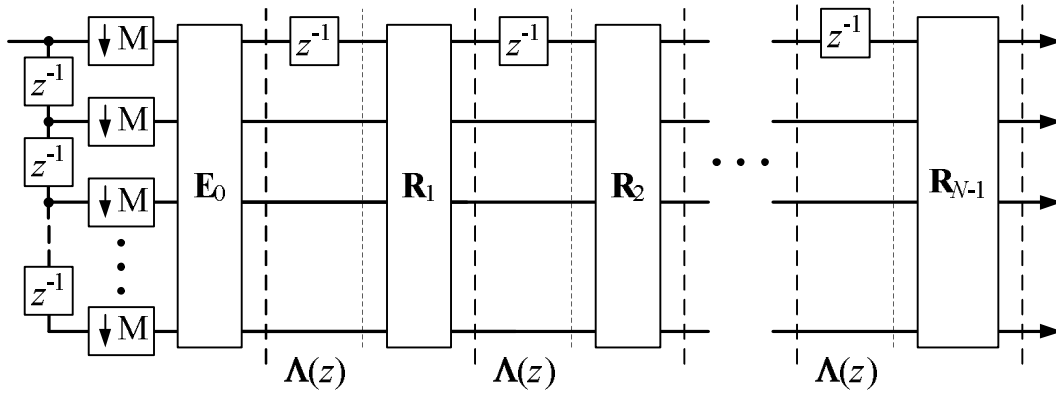


Рис. 3.1 – Решетчатая структура банка фильтров анализа

На практике широкое распространение получили банки фильтров с линейными фазо-частотными характеристиками (ФЧХ). Для ПУБФ анализа с линейной ФЧХ и четного числа каналов M хорошо известна факторизация передаточной полифазной матрицы $\mathbf{E}(z)$:

$$\mathbf{E}(z) = \mathbf{G}_{N-1}(z)\mathbf{G}_{N-2}(z) \dots \mathbf{G}_1(z)\mathbf{E}_0, \quad (3.2)$$

где

$$\mathbf{E}_0 = \frac{1}{\sqrt{2}} \mathbf{\Phi}_0 \mathbf{W} \text{diag}(\mathbf{I}_{M/2}, \mathbf{J}_{M/2}), \quad (3.3)$$

$$\mathbf{G}_i(z) = \frac{1}{2} \mathbf{\Phi}_0 \mathbf{W} \mathbf{\Lambda}(z) \mathbf{W}, i = 1, \dots, N-1, \quad (3.4)$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{I}_{M/2} & \mathbf{I}_{M/2} \\ \mathbf{I}_{M/2} & -\mathbf{I}_{M/2} \end{bmatrix}, \quad (3.5)$$

$$\mathbf{\Lambda}(z) = \text{diag}(\mathbf{I}_{M/2}, z^{-1}\mathbf{I}_{M/2}), \quad (3.6)$$

$$\mathbf{\Phi}_i = \text{diag}(\mathbf{U}_i, \mathbf{V}_i). \quad (3.7)$$

\mathbf{U}_i и \mathbf{V}_i произвольные ортогональные матрицы размера $M/2 \times M/2$. Кроме того, для упрощения обозначенной выше факторизации, для $i > 0$ матрица \mathbf{U}_i может быть представлена единичной матрицей [7], тогда

$$\Phi_i = \text{diag}(\mathbf{I}, \mathbf{V}_i), i > 0. \quad (3.8)$$

Решетчатая структура 4-канального банка фильтров анализа с линейной ФЧХ изображена на рис. 3.2 для $N = 2$.

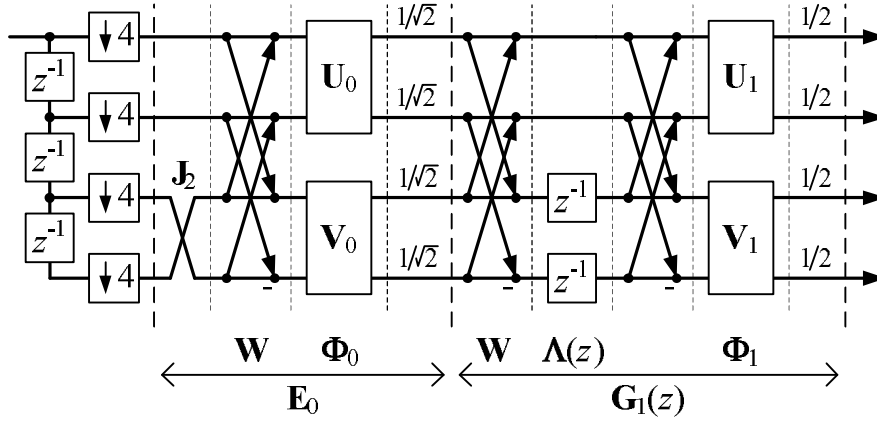


Рис. 3.2 – Решетчатая структура 4-канального банка фильтров анализа с линейной ФЧХ

Среди ПУБФ с линейной ФЧХ выделяются банки фильтров, у которых наблюдается попарная зеркальная симметрия частотных характеристик фильтров относительно $\pi/2$ [8]. Для этого передаточные функции или импульсные характеристики анализирующих фильтров должны удовлетворять соответственно следующим условиям:

$$H_{M-1-k}(z) = \pm H_k(-z) \text{ или } h_{M-1-k}(n) = \pm (-1)^n h_k(n), \quad (3.9)$$

где $k = 0, \dots, N-1$, $n = 0, \dots, L-1$, L – длина импульсной характеристики фильтра. Для четного числа каналов M ПУБФ с попарной зеркальной симметрией частотных характеристик фильтров может быть легко получен, если немного модифицировать факторизацию решетчатой структуры ПУБФ с линейной ФЧХ. Действительно, достаточно ассоциировать \mathbf{U}_i с \mathbf{V}_i в (3.7) и представить матрицу \mathbf{U}_i : следующим образом [8]:

$$\begin{aligned} \mathbf{U}_i &= \mathbf{\Gamma} \mathbf{V}_i \mathbf{\Gamma}, i = 0, \dots, N-2, \\ \mathbf{U}_{N-1} &= \mathbf{J}_{M/2} \mathbf{V}_{N-1} \mathbf{\Gamma}, \end{aligned} \quad (3.10)$$

где $\mathbf{J}_{M/2}$ – обратная единичная матрица; $\mathbf{\Gamma}$ – диагональная матрица, элементы которой определяются как $\gamma_{mm} = (-1)^{m-1}$, $m = 1, \dots, M/2$.

3.2 Регулярность и эффекты квантования коэффициентов

Для M -канального банка фильтров регулярность K -го порядка может быть определена как число нулей на частотах элайзинга $2\pi k/M$, $k = 1, \dots, M-1$ фильтра нижних частот $H_0(z)$. Чтобы получить K -ый порядок регулярности, полифазная матрица $\mathbf{E}(z)$ должна удовлетворять условию [9]: $\frac{d^n}{dz^n} \left\{ \mathbf{E}(z^M) [1 \ z^{-1} \ \dots \ z^{-(M-1)}]^T \right\}_{z=1} = c_n \mathbf{e}$, где $c_n \neq 0$ для $n = 0, \dots, K-1$, $\mathbf{e} =$

$[1\ 0\ 0\ 0]^T$. Например, в 8-ми канальном банке фильтров ($M = 8$) с регулярностью 1-го порядка ($K = 1$) значения амплитудно-частотных характеристик всех канальных фильтров $H_k(z)$, $k = 0, \dots, M - 1$ равны нулю на нулевой частоте. Таким образом, постоянная составляющая на выходе банка фильтров локализуется фильтром нижних частот $H_0(z)$ и не «просачивается» в другие канальные фильтры, т.е. не проявляется так называемый эффект *DC leakage*. В итоге, в обработанном изображении таким банком фильтров отсутствуют артефакты «шахматной доски» (checkerboard artifact).

К сожалению, описанные выше факторизации теряют свойство параунитарности, когда матрицы \mathbf{U}_i и \mathbf{V}_i представляются с ограниченной точностью [10]. Перфективная реконструкция не обеспечивается банком фильтров даже решетчатыми структурами с фиксированной запятой. Это обусловлено тем, что квантование компонент ортогональных матриц в общем случае приведет к изменению норм столбцов. Например, квантованная матрица оборотов Гивенса

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & Q(\cos \alpha) & -Q(\sin \alpha) & 0 \\ 0 & Q(\sin \alpha) & Q(\cos \alpha) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (3.11)$$

не ортогональна, так имеет две различные нормы столбцов: 1 и $\sqrt{Q^2(\cos \alpha) + Q^2(\sin \alpha)} \neq 1$, здесь Q - оператор квантования.

3.3 Кватернионы и ортогональные матрицы

3.3.1 Кватернионы

Кватернион – это гиперкомплексное число размерности 4 [11,12]

$$Q = q_1 + q_2i + q_3j + q_4k, \quad q_1, q_2, q_3, q_4 \in \mathbb{E}, \quad (3.12)$$

с одной реальной и тремя мнимыми частями. Мнимые единицы i, j, k соотносятся между собой согласно следующим уравнениям:

$$i^2 = j^2 = k^2 = ijk = -1; \quad ij = -ji = k; \quad jk = -kj = i. \quad (3.13)$$

Особенность умножения кватернионов в том, что оно не подчиняется переместительному закону умножения ($PQ \neq QP$, где P, Q – произвольные кватернионы), но подчиняется сочетательному. Сложение кватернионов осуществляется аналогично сложению комплексных чисел. Сопряженным для кватерниона вида (3.12) будет кватернион:

$$\bar{Q} = q_1 - q_2i - q_3j - q_4k. \quad (3.14)$$

Модуль кватерниона определяется аналогично модулю комплексного числа:

$$|Q| = \sqrt{q_1^2 + q_2^2 + q_3^2 + q_4^2}. \quad (3.15)$$

Основные свойства модуля кватерниона также совпадают с подобными свойствами модуля комплексного числа, в частности:

$$Q \cdot \bar{Q} = |Q|^2; |Q_1 \cdot Q_2|^2 = |Q_1|^2 \cdot |Q_2|^2. \quad (3.16)$$

Деление кватернионов определяется как умножение на обратную величину $Q^{-1} = \bar{Q}/|Q|^2$. При этом удовлетворяется тождество $Q \cdot Q^{-1} = Q^{-1} \cdot Q = 1$. Полярная форма кватерниона определяется следующим образом [11]:

$$\begin{cases} q_1 = |Q| \cdot \cos \varphi \\ q_2 = |Q| \cdot \sin \varphi \cdot \cos \psi \\ q_3 = |Q| \cdot \sin \varphi \cdot \cos \psi \cdot \cos \chi' \\ q_4 = |Q| \cdot \sin \varphi \cdot \cos \psi \cdot \sin \chi \end{cases} \quad (3.17)$$

где φ, ψ, χ – произвольные значения углов ($0 \leq \varphi \leq \pi, 0 \leq \psi \leq \pi, 0 \leq \chi \leq 2\pi$). Полярное представление кватернионов сокращает число степеней свободы до трех и позволяет легко параметризовать кватернионы с фиксированными модулями, в частности, с единичными кватернионами, модуль которых равен 1.

3.3.2 Матрицы умножения

Операции сложения и умножения кватернионов также могут быть реализованы с использованием векторно-матричной арифметики. Для этого кватернионы представляются в виде четырехмерных векторов: $P = [p_1 \ p_2 \ p_3 \ p_4]^T, Q = [q_1 \ q_2 \ q_3 \ q_4]^T$. В этом случае сложение кватернионов будет производиться по правилу сложения векторов, а умножение кватернионов в векторной форме производится по правилу умножения вектора на матрицу:

$$\begin{aligned} R = P \cdot Q &= \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{bmatrix} = \begin{bmatrix} p_1 & -p_2 & -p_3 & -p_4 \\ p_2 & p_1 & -p_4 & p_3 \\ p_3 & p_4 & p_1 & -p_2 \\ p_4 & -p_3 & p_2 & p_1 \end{bmatrix} \times \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} = \begin{bmatrix} p_1q_1 - p_2q_2 - p_3q_3 - p_4q_4 \\ p_2q_1 + p_1q_2 - p_4q_3 + p_3q_4 \\ p_3q_1 + p_4q_2 + p_1q_3 - p_2q_4 \\ p_4q_1 - p_3q_2 + p_2q_3 + p_1q_4 \end{bmatrix} \\ &= \begin{bmatrix} q_1p_1 - q_2p_2 - q_3p_3 - q_4p_4 \\ q_2p_1 + q_1p_2 + q_4p_3 - q_3p_4 \\ q_3p_1 - q_4p_2 + q_1p_3 + q_2p_4 \\ q_4p_1 + q_3p_2 - q_2p_3 + q_1p_4 \end{bmatrix} = \begin{bmatrix} q_1 & -q_2 & -q_3 & -q_4 \\ q_2 & q_1 & q_4 & -q_3 \\ q_3 & -q_4 & q_1 & q_2 \\ q_4 & q_3 & -q_2 & q_1 \end{bmatrix} \times \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix}. \end{aligned} \quad (3.18)$$

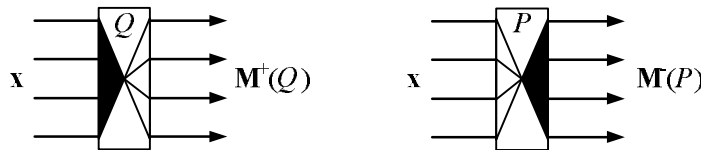


Рис. 3.3 – Графические изображения операндов умножения кватернионов

Данные выражения можно представить в более простом виде:

$$R = P \cdot Q = \mathbf{M}^+(P) \cdot Q = \mathbf{M}^-(Q) \cdot P, \quad (3.19)$$

где матрица $\mathbf{M}^+(P)$ – левый операнд умножения кватерниона P , а матрица $\mathbf{M}^-(Q)$ представляет собой правый операнд умножения кватерниона Q :

$$\mathbf{M}^+(P) = \begin{bmatrix} p_1 & -p_2 & -p_3 & -p_4 \\ p_2 & p_1 & -p_4 & p_3 \\ p_3 & p_4 & p_1 & -p_2 \\ p_4 & -p_3 & p_2 & p_1 \end{bmatrix}; \mathbf{M}^-(Q) = \begin{bmatrix} q_1 & -q_2 & -q_3 & -q_4 \\ q_2 & q_1 & q_4 & -q_3 \\ q_3 & -q_4 & q_1 & q_2 \\ q_4 & q_3 & -q_2 & q_1 \end{bmatrix}. \quad (3.20)$$

Для изображения операндов умножения кватернионов на структурных схемах могут быть использованы следующие графические обозначения, как показано на рис. 3.3.

```
function M = quatmulmtxr(x);
% construction of right-multiplication matrix for hypercomplex number
% x hypercomplex number
% M multiplication matrix

M = [ x(1) -x(2) -x(3) -x(4); ...
      x(2) x(1) x(4) -x(3); ...
      x(3) -x(4) x(1) x(2); ...
      x(4) x(3) -x(2) x(1) ];
```

Обе матрицы операндов умножения являются ортогональными, то есть справедливо условие:

$$\mathbf{M}^{\pm}(Q)^{-1} = \mathbf{M}^{\pm}(Q)^T, \quad (3.21)$$

и имеют детерминант равный +1, что позволяет отнести их к группе специальных матриц размерностью 4×4 , называемых $SO(4)$ [13,14]. Для рассматриваемых матриц операторов умножения верны следующие равенства:

$$\mathbf{M}^+(Q_{N-1})\mathbf{M}^+(Q_{N-2}) \dots \mathbf{M}^+(Q_0) = \mathbf{M}^+(Q_{N-1} Q_{N-2} \dots Q_0), \quad (3.22)$$

$$\mathbf{M}^-(Q_{N-1})\mathbf{M}^-(Q_{N-2}) \dots \mathbf{M}^-(Q_0) = \mathbf{M}^-(Q_{N-1} Q_{N-2} \dots Q_0). \quad (3.23)$$

Существует другое интересное и полезное соотношение:

$$\mathbf{M}^{\pm}(\bar{Q}) = \mathbf{M}^{\pm}(Q)^T. \quad (3.24)$$

Соотношения между матрицами умножения кватернионов и произвольными 4×4 и 8×8 ортогональными матрицами является интригующим с точки зрения перспектив синтеза ПУБФ.

3.3.3 Параметризация ортогональных матриц

Для обоснования применения умножителей кватернионов в качестве составляющих блоков ПУБФ было выдвинуто и доказано несколько теорем, основными из которых являются следующие:

1) для каждой ортогональной матрицы \mathbf{A} размерностью 4×4 существует уникальная (с точностью до знака) пара единичных кватернионов P и Q , таких что [14]:

$$\mathbf{A} = \mathbf{M}^+(P) \cdot \mathbf{M}^-(Q) = \mathbf{M}^-(Q) \cdot \mathbf{M}^+(P); \quad (3.25)$$

2) когда блочно-диагональная матрица $\text{diag}(\mathbf{U}, \mathbf{V})$ размерностью 8×8 образована из двух произвольных ортогональных матриц \mathbf{U} и \mathbf{V} размерностью 4×4 , то существуют четыре единичных кватерниона P, Q, R, S , для которых справедливо соотношение [15,16]:

$$\text{diag}(\mathbf{U}, \mathbf{V}) = \text{diag}(\mathbf{M}^-(S), \mathbf{M}^-(S)) \cdot \text{diag}(\mathbf{M}^-(\bar{R}), \mathbf{M}^-(R)) \cdot \text{diag}(\mathbf{M}^+(\bar{Q}), \mathbf{M}^+(Q)) \cdot \text{diag}(\mathbf{M}^+(P), \mathbf{M}^+(P)). \quad (3.26)$$

В случае, когда оба блока диагональной матрицы равны ($\mathbf{U} = \mathbf{V}$) выражение (3.26) примет следующий вид:

$$\text{diag}(\mathbf{V}, \mathbf{V}) = \text{diag}(\mathbf{M}^-(Q), \mathbf{M}^-(Q)) \cdot \text{diag}(\mathbf{M}^+(P), \mathbf{M}^+(P)). \quad (3.27)$$

Если же блок \mathbf{U} диагональной матрицы $\text{diag}(\mathbf{U}, \mathbf{V})$ является единичной матрицей, то выражение (3.26) преобразуется к виду:

$$\text{diag}(\mathbf{I}_4, \mathbf{V}) = \text{diag}(\mathbf{M}^-(Q), \mathbf{M}^-(Q)) \cdot \text{diag}(\mathbf{M}^-(\bar{Q}), \mathbf{M}^-(Q)) \cdot \text{diag}(\mathbf{M}^+(\bar{P}), \mathbf{M}^+(P)) \cdot \text{diag}(\mathbf{M}^+(P), \mathbf{M}^+(P)). \quad (3.28)$$

3.3.4 Умножитель кватернионов как составной блок ПУБФ

Факторизация (3.25) всегда соответствуют ортогональному преобразованию сигналов, даже если ее компоненты квантованы. Это происходит потому, что столбцы каждой матрицы умножения кватернионов составлены из одних и тех же элементов с точностью до знака, например:

$$\begin{bmatrix} Q(q_1) & -Q(q_2) & -Q(q_3) & -Q(q_4) \\ Q(q_2) & Q(q_1) & -Q(q_4) & Q(q_3) \\ Q(q_3) & Q(q_4) & Q(q_1) & -Q(q_2) \\ Q(q_4) & -Q(q_3) & Q(q_2) & Q(q_1) \end{bmatrix}. \quad (3.29)$$

Видно, что как строчные так и столбцовые нормы постоянны и равны $\sqrt{Q(q_1)^2 + Q(q_2)^2 + Q(q_3)^2 + Q(q_4)^2}$, но не обязательно единичны, невзирая на конечную точность представления матриц (или кватернионных коэффициентов). Таким образом, матрица \mathbf{A} в (3.25) всегда представляет ортогональное преобразование, а умножение кватернионов может рассматриваться как основной строительный блок ПУБФ.

Произвольная 4×4 матрицы требует хранения 16 действительных чисел и 32 чисел для блочно-диагональной матрицы 8×8 . В данном же случае, необходимо запомнить в памяти 8 и 16 действительных чисел (2 и 4 кватерниона) соответственно. Один из недостатков кватернионной факторизации заключается в большом количестве вычислений, что является ценой за сохранение ортогональности. Вычислительная сложность факторизаций (3.25) в два раза, а (3.26) в четыре раза больше, чем простое умножение матрицы 4×4 на вектор. Однако, учитывая специфическую структуру матриц умножения кватернионов, можно выполнить данные операции только за 8 реальных умножений [17]. Реализация умножителей кватернионов на основе распределенной арифметики [18] или с использованием 4-х мерных алгоритмов CORDIC [19] хорошо согласуется со структурами FPGA.

3.4 Решетчатые структуры ПУБФ на основе кватернионов

3.4.1 Четырех канальный ПУБФ

Матрицы \mathbf{E}_0 и \mathbf{R}_i в факторизации (3.1) решетчатой структуры общего параунитарного банка фильтров анализа, как показано в [20], можно выразить через матрицы умножения кватернионов следующим образом:

$$\mathbf{E}_0 = \mathbf{M}^+(Q_0) \cdot \mathbf{M}^-(P_0), \quad (3.30)$$

$$\mathbf{R}_i = \mathbf{M}^\pm(Q_i), i = 1, \dots, N - 1, \quad (3.31)$$

где P_0 и все Q_i – некоторые единичные кватернионы, тогда передаточная функция решетчатой структуры общего ПУБФ анализа будет равна:

$$\mathbf{E}(z) = \mathbf{M}^\pm(Q_{N-1})\Lambda(z)\mathbf{M}^\pm(Q_{N-2})\Lambda(z) \dots \mathbf{M}^\pm(Q_1)\Lambda(z)\mathbf{M}^+(Q_0)\mathbf{M}^-(P_0). \quad (3.32)$$

Как доказано в [20], если в (3.31) для всех $i = 1, \dots, N - 1$ используются левые операнды умножения и значения кватернионов удовлетворяют равенству:

$$P_0 = \pm \frac{1}{2} \bar{O} \cdot \overline{Q_{N-1} \cdot Q_{N-2} \cdot \dots \cdot Q_0}, \quad (3.33)$$

где O – кватернион, все части которого равны 1: $O = 1 + i + j + k$, то для решетчатой структуры ПУБФ (3.32) выполняется условие регулярности 1-го порядка.

Решетчатая структура 4-канального общего ПУБФ анализа (для $N = 3$), определяемая выражением (3.32), показана на рис. 3.4.

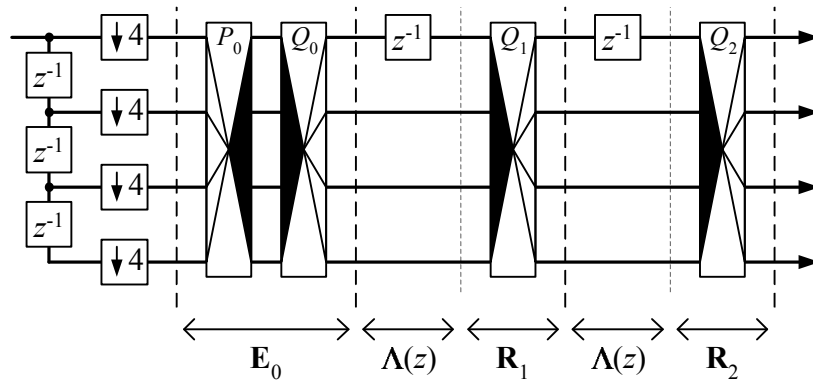


Рис. 3.4 – Решетчатая структура 4-канального ПУБФ анализа на кватернионах

На основе связи между матрицами передаточных функций фильтров банков анализа и синтеза для обеспечения перфективной реконструкции требуемый банк фильтров синтеза может быть определен как:

$$\mathbf{D}(z) = \mathbf{E}(z)^{-1} c z^{-\Delta} \mathbf{I}. \quad (3.34)$$

Подставив (3.32) в формулу (3.34) и учитывая соотношения (3.21) и (3.24) для матриц операндов умножения кватернионов, можно получить следующую факторизацию:

$$\mathbf{D}(z) = c z^{-\Delta} \mathbf{M}^-(\overline{P_0}) \mathbf{M}^+(\overline{Q_0}) \Lambda(z)^{-1} \mathbf{M}^\pm(\overline{Q_1}) \dots \Lambda(z)^{-1} \mathbf{M}^\pm(\overline{Q_{N-1}}) \Lambda(z)^{-1} \mathbf{M}^\pm(\overline{Q_N}), \quad (3.35)$$

где \bar{Q}_i, \bar{P}_0 – кватернионы, комплексно-сопряженные к Q_i, P_0 . Приняв константу c равной 1 (для сохранения масштаба) и используя следующие замены:

$$\hat{\Lambda}(z) = \Lambda(z)^{-1} z^{-1} \mathbf{I}_2, \quad (3.36)$$

$$\hat{\mathbf{E}}_0 = \mathbf{M}^-(\bar{P}_0) \mathbf{M}^+(\bar{Q}_0), \quad (3.37)$$

$$\hat{\mathbf{R}}_i = \mathbf{M}^\pm(\bar{Q}_i), i = 1, \dots, N - 1, \quad (3.38)$$

окончательная факторизация для ПУБФ синтеза будет выглядеть следующим образом:

$$\begin{aligned} \mathbf{D}(z) &= \mathbf{M}^-(\bar{P}_0) \mathbf{M}^+(\bar{Q}_0) \hat{\Lambda}(z) \mathbf{M}^\pm(\bar{Q}_1) \dots \hat{\Lambda}(z) \mathbf{M}^\pm(\bar{Q}_{N-2}) \hat{\Lambda}(z) \mathbf{M}^\pm(\bar{Q}_{N-1}) \\ &= \hat{\mathbf{E}}_0 \hat{\Lambda}(z) \hat{\mathbf{R}}_1 \dots \hat{\Lambda}(z) \hat{\mathbf{R}}_{N-2} \hat{\Lambda}(z) \hat{\mathbf{R}}_{N-1}. \end{aligned} \quad (3.39)$$

Факторизация (3.39) описывает решетчатую структуру ПУБФ синтеза (для $N = 3$), показанную на рис. 3.5.

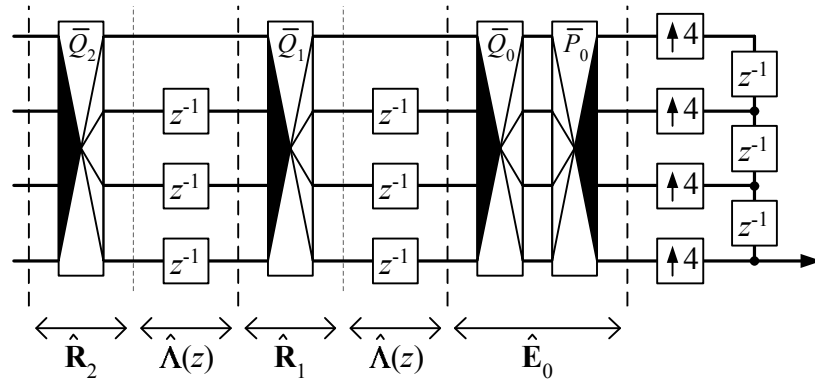


Рис. 3.5 – Решетчатая структура 4-канального банка фильтров синтеза на кватернионах

Реализация параунитарного банка фильтров анализа с линейной ФЧХ (3.2) на основе умножителей кватернионов предполагает определение Φ_0 и Φ_i как [21]:

$$\Phi_0 = \mathbf{M}^-(P_0) \mathbf{M}^+(Q_0), \quad (3.40)$$

$$\Phi_i = \mathbf{M}^-(P_i), i = 1, \dots, N - 1, \quad (3.41)$$

где все P_i и Q_0 – некоторые единичные кватернионы. Если в (3.40) и (3.41)

$$Q_0 = \pm \frac{1}{\sqrt{2}} \overline{P_0 \cdot P_1 \cdot \dots \cdot P_{N-1}} \cdot \bar{A}, \quad (3.42)$$

то факторизация (3.2), (3.40) – (3.42) соответствует регулярности 1-го порядка. Здесь A – специальный кватернион: $A = 1 + i$.

Решетчатая структура ПУБФ анализа с линейной ФЧХ на основе умножителей кватернионов (для $N = 3$) представлена на рис. 3.6.

Факторизация ПУБФ синтеза на кватернионах задается следующим образом:

$$\mathbf{D}(z) = \hat{\mathbf{E}}_0 \hat{\mathbf{G}}_1 \dots \hat{\mathbf{G}}_{N-2} \hat{\mathbf{G}}_{N-1}, \quad (3.43)$$

$$\hat{\mathbf{E}}_0 = \text{diag}(\mathbf{I}_2, \mathbf{J}_2) \mathbf{W} \hat{\Phi}_0 \frac{1}{\sqrt{2}}, \quad (3.44)$$

$$\hat{\mathbf{G}}_i(z) = \mathbf{W} \hat{\Lambda}(z) \mathbf{W} \hat{\Phi}_0 \frac{1}{2}, \quad i = 1, \dots, N - 1, \quad (3.45)$$

$$\hat{\Lambda}(z) = \text{diag}(z^{-1} \mathbf{I}_2, \mathbf{J}_2), \quad (3.46)$$

$$\hat{\Phi}_0 = \mathbf{M}^+(\overline{Q}_0) \mathbf{M}^-(\overline{P}_0), \quad (3.47)$$

$$\hat{\Phi}_i = \mathbf{M}^-(\overline{P}_i), \quad i = 1, \dots, N - 1. \quad (3.48)$$

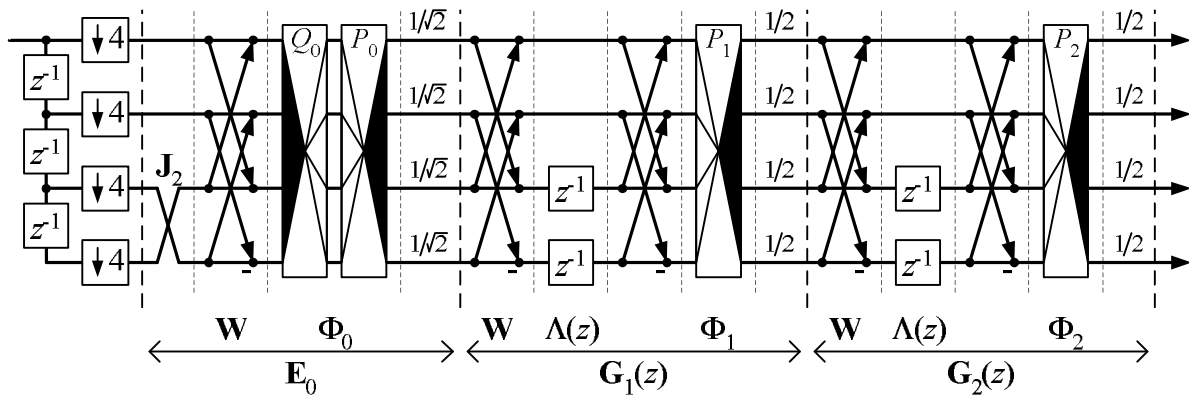


Рис. 3.6 – Решетчатая структура 4-канального банка фильтров анализа с линейной ФЧХ на квартернионах

Решетчатая структура 4-канального ПУБФ синтеза с линейной ФЧХ, соответствующая факторизации (3.43) – (3.48) (для $N = 3$), изображена на рис. 3.7.

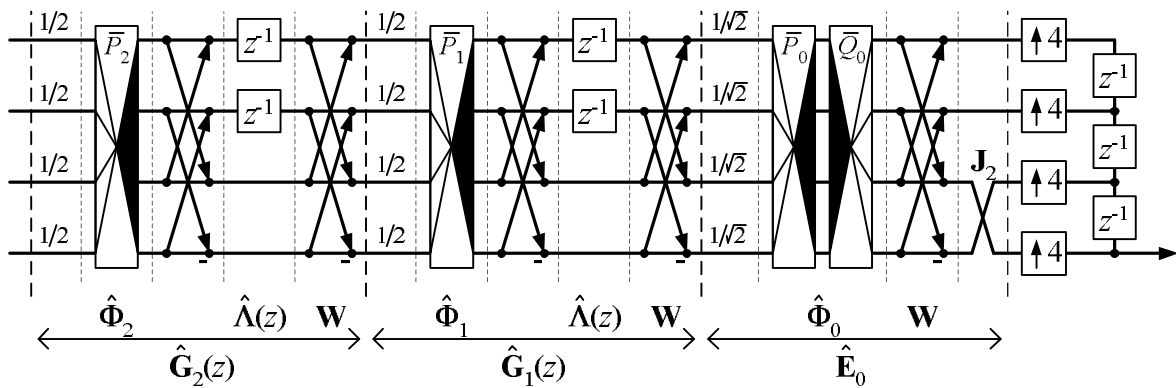


Рис. 3.7 – Решетчатая структура 4-канального банка фильтров синтеза с линейной ФЧХ на квартернионах

Решетчатая структура на квартернионах ПУБФ с линейной ФЧХ и попарно зеркальной симметрией частотных характеристик фильтров получается в результате факторизации (3.2), (3.10) для соответственно определенных матриц Φ_i и Φ_{N-1} [20]:

$$\Phi_i = \mathbf{M}^-(\overline{P}_i), \quad i = 1, \dots, N - 2, \quad (3.49)$$

$$\Phi_{N-1} = \mathbf{M}^{-1}(P_{N-1})\text{diag}(\mathbf{J}_2\Gamma, \mathbf{I}_2), \quad (3.50)$$

где $\Gamma = \text{diag}(1, -1)$ и P_i – некоторые единичные кватернионы, выбор которых в соответствии с

$$P_{N-1} = \pm \frac{1}{\sqrt{2}} A \cdot \overline{P_0 \cdot P_1 \cdot \dots \cdot P_{N-2}}, \quad (3.51)$$

обеспечивает регулярность первого порядка банку фильтров.

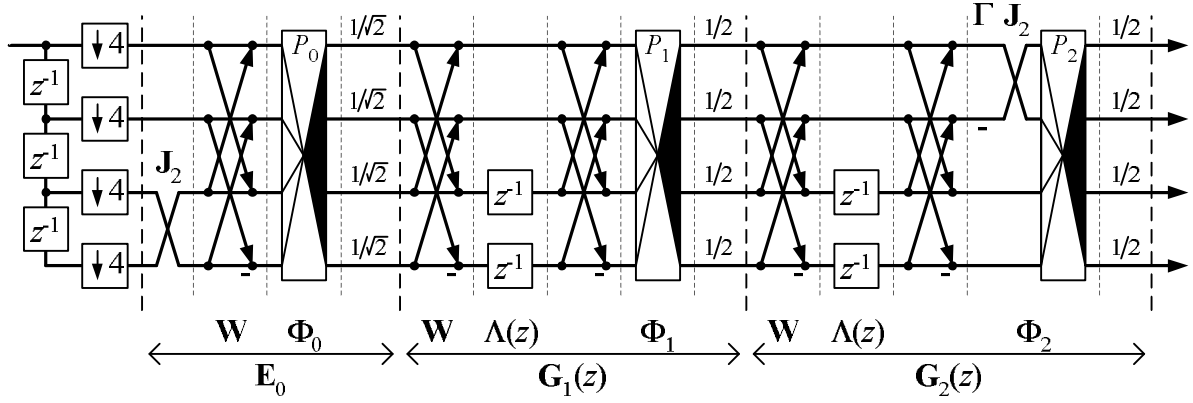


Рис. 3.8 – Решетчатая структура 4-канального банка фильтров анализа с попарно зеркальной симметрией частотных характеристик фильтров на кватернионах

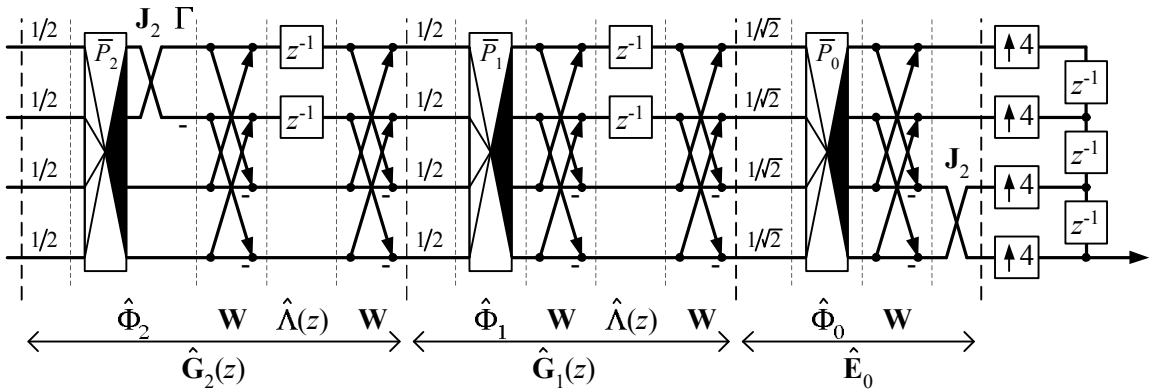


Рис. 3.9 – Решетчатая структура 4-канального банка фильтров синтеза с попарно зеркальной симметрией частотных характеристик фильтров на кватернионах

Соответствующая факторизация передаточной функции банка фильтров синтеза задается следующим образом:

$$\mathbf{D}(z) = \hat{\mathbf{E}}_0 \hat{\mathbf{G}}_1 \dots \hat{\mathbf{G}}_{N-2} \hat{\mathbf{G}}_{N-1}, \quad (3.52)$$

$$\hat{\mathbf{E}}_0 = \text{diag}(\mathbf{I}_2, \mathbf{J}_2) \mathbf{W} \hat{\Phi}_0 \frac{1}{\sqrt{2}}, \quad (3.53)$$

$$\hat{\mathbf{G}}_i(z) = \mathbf{W} \hat{\Lambda}(z) \mathbf{W} \hat{\Phi}_i \frac{1}{2}, \quad i = 1, \dots, N-1, \quad (3.54)$$

$$\hat{\Lambda}(z) = \text{diag}(z^{-1} \mathbf{I}_2, \mathbf{J}_2), \quad (3.55)$$

$$\widehat{\Phi}_{N-1} = \text{diag}(\mathbf{\Gamma}_2, \mathbf{I}_2) \mathbf{M}^{-}(\overline{P_{N-1}}), \quad (3.56)$$

$$\widehat{\Phi}_i = \mathbf{M}^{-}(\overline{P_i}), i = 1, \dots, N - 2. \quad (3.57)$$

Решетчатые структуры ПУБФ анализа и синтеза с попарно зеркальной симметрией частотных характеристик фильтров приведены на рис. 3.8 и рис. 3.9 соответственно ($N = 3$).

3.4.2 Восьми канальный ПУБФ

Вариант кватернионной факторизации (3.2) – (3.10) для 8-ми канального ПУБФ с линейной ФЧХ анализа определяется следующими подстановками [5,15,22,23]:

$$\Phi_i = \text{diag}(\mathbf{M}^{-}(\overline{R_i}), \mathbf{M}^{-}(R_i)) \cdot \text{diag}(\mathbf{M}^{+}(\overline{Q_i}), \mathbf{M}^{+}(Q_i)), i = 1, \dots, N - 1, \quad (3.58)$$

$$\Phi_0 = \text{diag}(\mathbf{M}^{-}(S_0), \mathbf{M}^{-}(S_0)) \cdot \text{diag}(\mathbf{M}^{-}(\overline{R_0}), \mathbf{M}^{-}(R_0)) \cdot \text{diag}(\mathbf{M}^{+}(\overline{Q_0}), \mathbf{M}^{+}(Q_0)) \cdot \text{diag}(\mathbf{M}^{+}(P_0), \mathbf{M}^{+}(P_0)), \quad (3.59)$$

где S_0, P_0 и все Q_i, R_i – единичные кватернионы. Данный ПУБФ (рис. 3.10) будет банком фильтров с регулярностью первого порядка, если

$$R_{N-1} = \frac{1}{2} \overline{Q_{N-1}} \cdot \dots \cdot \overline{Q_0} \cdot P_0 \cdot O \cdot \overline{R_0} \cdot S_0 \cdot \overline{R_1} \cdot \dots \cdot \overline{R_{N-2}}. \quad (3.60)$$

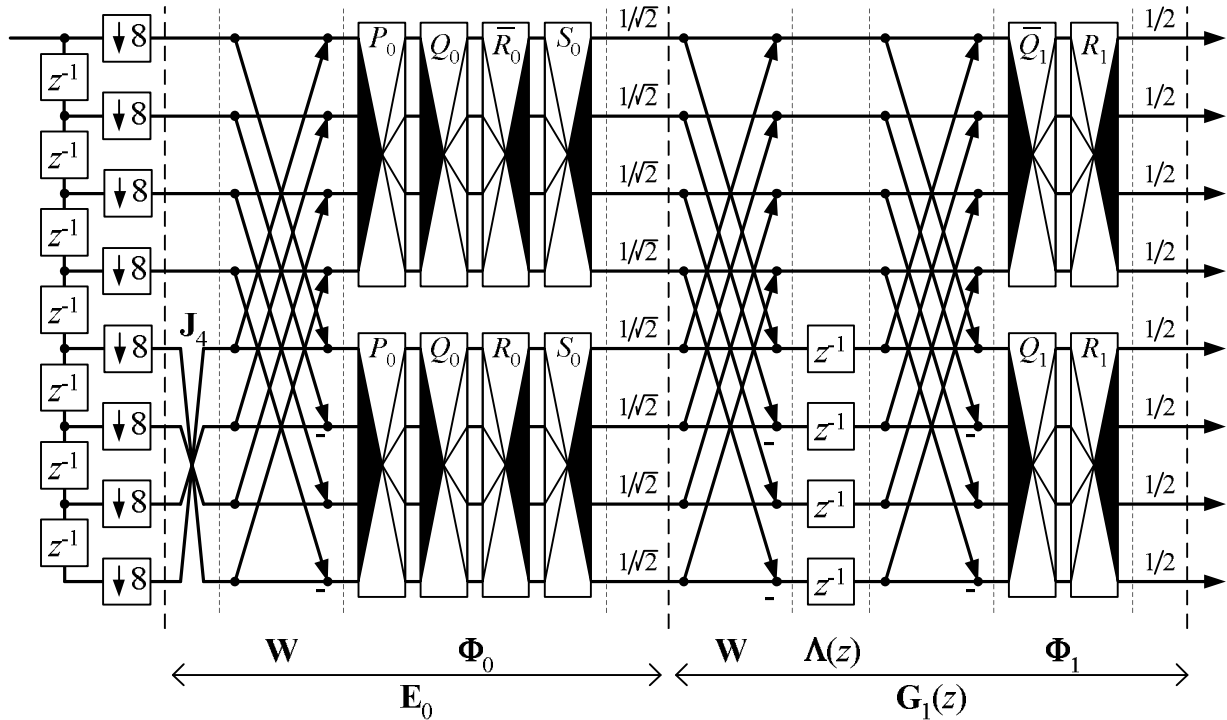


Рис. 3.10 – Решетчатая структура 8-канального банка фильтров анализа с линейной ФЧХ на кватернионах (для $N = 2$)

Для соответствующего 8-канального банка синтеза (рис. 3.11) в факторизацию (3.43) – (3.48) вводятся следующие подстановки:

$$\widehat{\Phi}_i = \text{diag}(\mathbf{M}^{+}(Q_i), \mathbf{M}^{+}(\overline{Q_i})) \cdot \text{diag}(\mathbf{M}^{-}(R_i), \mathbf{M}^{-}(\overline{R_i})), i = 1, \dots, N - 1, \quad (3.61)$$

$$\hat{\Phi}_0 = \text{diag}(\mathbf{M}^+(\bar{P}_0), \mathbf{M}^+(\bar{P}_0)) \cdot \text{diag}(\mathbf{M}^+(Q_0), \mathbf{M}^+(Q_0)) \cdot \text{diag}(\mathbf{M}^-(R_0), \mathbf{M}^-(R_0)) \cdot \text{diag}(\mathbf{M}^-(\bar{S}_0), \mathbf{M}^-(\bar{S}_0)), \quad (3.62)$$

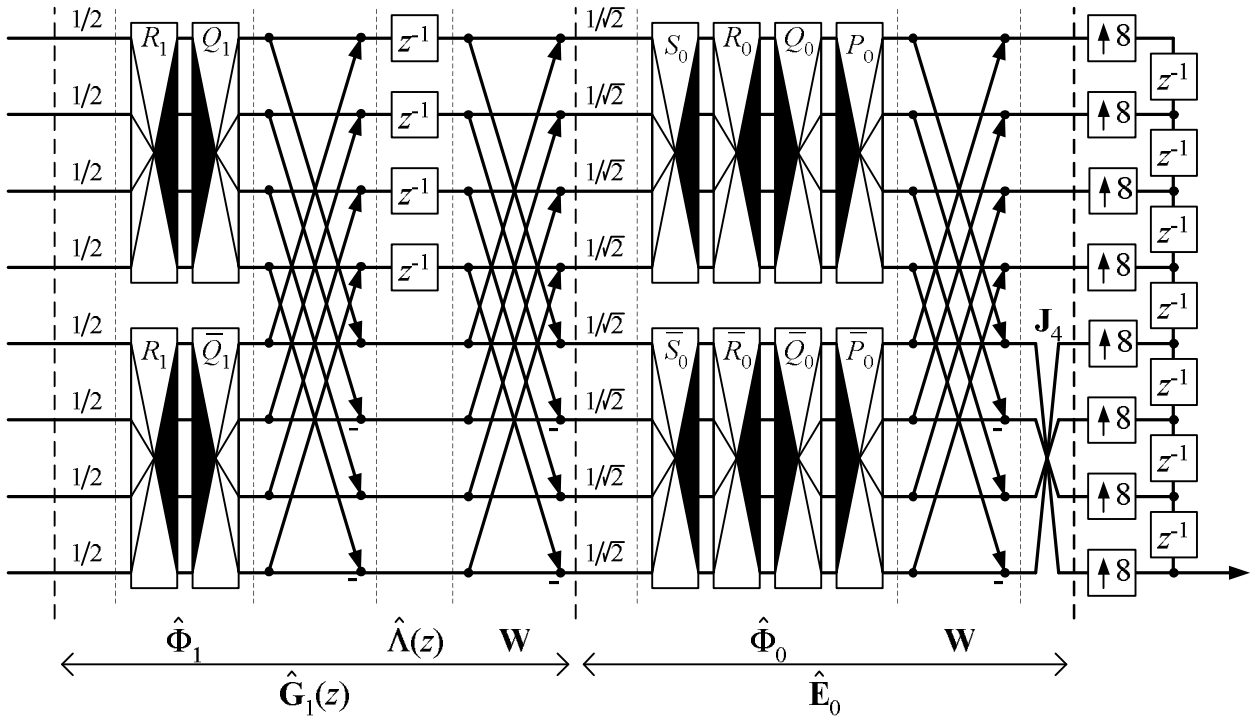


Рис. 3.11 – Решетчатая структура 8-канального банка фильтров синтеза с линейной ФЧХ на кватернионах (для $N = 2$)

Решетчатые структуры 8-ми канальных ПУБФ с линейной ФЧХ и попарно зеркальной симметрией частотных характеристик фильтров (3.2) – (3.10) на квантаренионах получается для соответственно определенных матриц Φ_i и Φ_{N-1} [22,23]:

для банка анализа (рис. 3.12):

$$\Phi_i = \text{diag}(\Gamma, \mathbf{I}_4) \cdot \text{diag}(\mathbf{M}^-(Q_i), \mathbf{M}^-(Q_i)) \cdot \text{diag}(\mathbf{M}^+(P_i), \mathbf{M}^+(P_i)) \cdot \text{diag}(\Gamma, \mathbf{I}_4), i = 0, \dots, N - 1, \quad (3.63)$$

$$\Phi_{N-1} = \text{diag}(\mathbf{J}_4, \mathbf{I}_4) \cdot \text{diag}(\mathbf{M}^-(Q_{N-1}), \mathbf{M}^-(Q_{N-1})) \cdot \text{diag}(\mathbf{M}^+(P_{N-1}), \mathbf{M}^+(P_{N-1})) \cdot \text{diag}(\Gamma, \mathbf{I}_4), \quad (3.64)$$

где все Q_i, P_i – единичные кватернионы, выбор которых в соответствии с соотношением

$$P_{N-1} = \frac{1}{2} \hat{E} \bar{Q}_{N-1} \cdot \dots \cdot \bar{Q}_0 \cdot P_0 \cdot \bar{O} \cdot \bar{P}_0 \cdot \bar{P}_1 \cdot \dots \cdot \bar{P}_{N-2}, \quad (3.65)$$

обеспечивает регулярность первого порядка банку фильтров;

для банка синтеза (рис. 3.13):

$$\hat{\Phi}_i = \text{diag}(\Gamma, \mathbf{I}_4) \cdot \text{diag}(\mathbf{M}^+(\bar{P}_i), \mathbf{M}^+(\bar{P}_i)) \cdot \text{diag}(\mathbf{M}^-(\bar{Q}_i), \mathbf{M}^-(\bar{Q}_i)) \cdot \text{diag}(\Gamma, \mathbf{I}_4), i = 0, \dots, N - 2, \quad (3.66)$$

$$\hat{\Phi}_{N-1} = \text{diag}(\Gamma, \mathbf{I}_4) \cdot \text{diag}(\mathbf{M}^+(\bar{P}_{N-1}), \mathbf{M}^+(\bar{P}_{N-1})) \quad (3.67)$$

$$\cdot \text{diag}(\mathbf{M}^-(\bar{Q}_{N-1}), \mathbf{M}^-(\bar{Q}_{N-1})) \cdot \text{diag}(\mathbf{J}_4, \mathbf{I}_4),$$

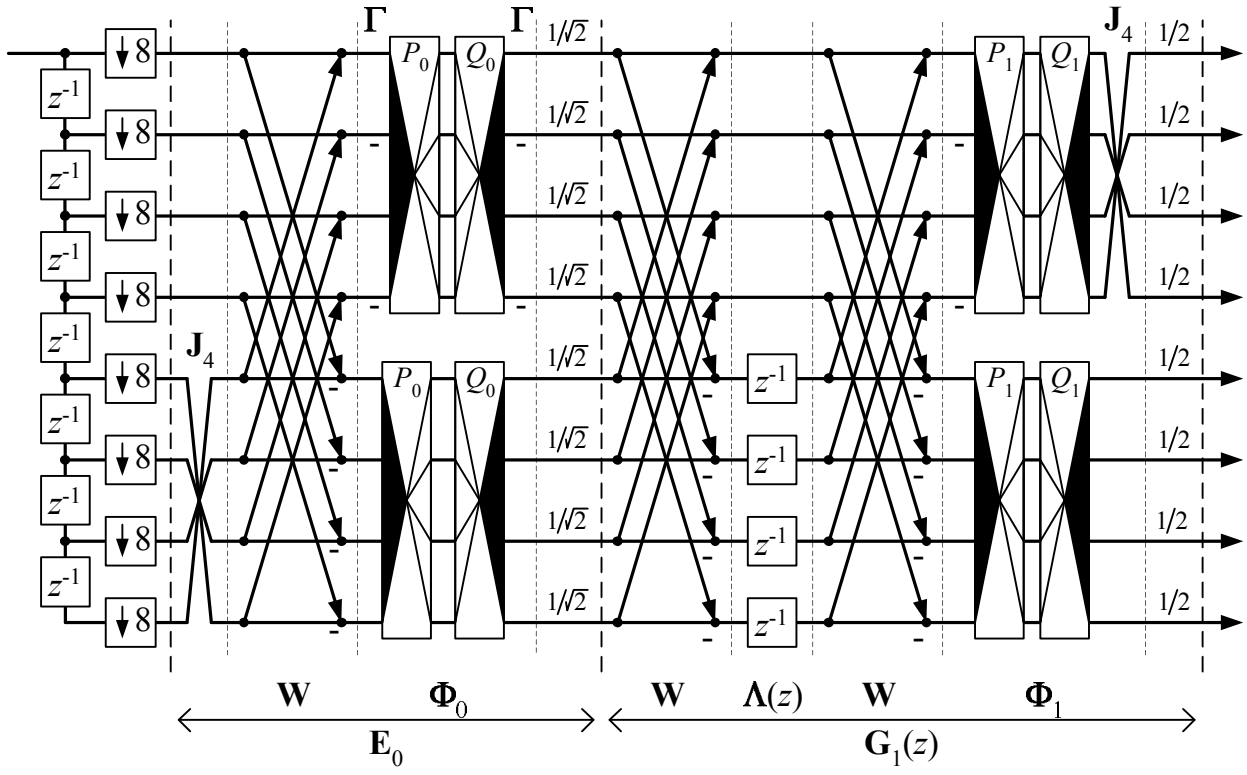


Рис. 3.12 – Решетчатая структура 8-канального банка фильтров анализа с попарно зеркальной симметрией частотных характеристик фильтров на кватернионах (для $N = 2$)

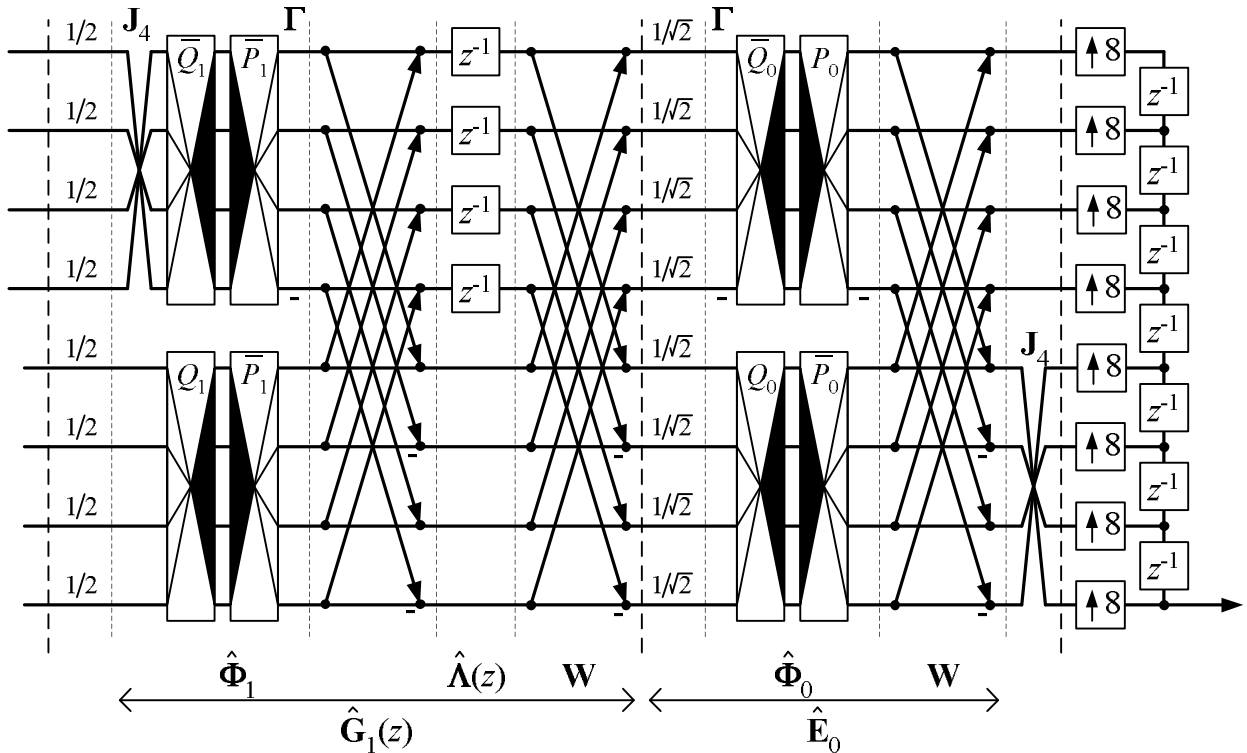


Рис. 3.13 – Решетчатая структура 8-канального банка фильтров синтеза с попарно зеркальной симметрией частотных характеристик фильтров на кватернионах (для $N = 2$)

3.5 Примеры синтеза

3.5.1 Процедура расчета

Для получения ПУБФ с высоким коэффициентом эффективности кодирования (coding gain) и хорошей частотной избирательностью кватернионные коэффициенты оптимизируются путем минимизации двух показателей:

- степень ослабления в полосе задержки каналов, выраженная в терминах энергии:

$$\varepsilon_{SBE} = \sum_{k=0}^{M-1} \int_{\omega \in \Omega_k} |H_k(e^{j\omega})|^2 d\omega, \quad (3.68)$$

где Ω_k определяют полосу задержки k -го фильтра банка фильтров;

- коэффициент эффективности кодирования:

$$CG = 10 \lg \frac{\frac{1}{M} \sum_{k=0}^{M-1} \sigma_{x_k}^2}{\left(\prod_{k=0}^{M-1} \sigma_{x_k}^2 \right)^{\frac{1}{M}}}, \quad (3.69)$$

где $\sigma_{x_k}^2$ - дисперсия сигнала в k -ом канале банка фильтров, значение которой для всех M каналов банка фильтров определяются через диагональные элементы автокорреляционной матрицы \mathbf{R}_{yy} выходного сигнала $y(n)$: $\sigma_{x_k}^2 = [\mathbf{R}_{yy}]_{kk}$. Автокорреляционная матрица $\mathbf{R}_{yy} = \mathbf{H}\mathbf{R}_{xx}\mathbf{H}^T$. Здесь \mathbf{R}_{xx} - автокорреляционная матрица входного сигнала $x(n)$, а \mathbf{H} - матрица преобразования, сформированная из импульсных характеристик фильтров банка следующим образом: $[\mathbf{H}]_{kn} = h_k(L-1-n)$, где $k = 0, \dots, M-1$, $n = 0, \dots, L-1$. В рассматриваемых ниже примерах синтеза ПУБФ матрица \mathbf{R}_{xx} конструировалась для входного сигнала $x(n)$, который формировался с помощью авторегрессионной модели первого порядка $AR(1)$ с единичной дисперсией и коэффициентом корреляции 0,95. Такая модель сигнала соответствует только обработке естественных изображений, для других типов сигналов потребуются другие модели.

Кватернионные коэффициенты решетчатых структур ПУБФ, приведенных выше, для процедуры синтеза представляются в полярной форме (3.17). Для решения задачи оптимизации (минимизация ε_{SBE} и CG) используются программы `fminsearch` и `fminunc` из пакета MATLAB. Далее, полученные коэффициенты конвертируются в рациональные числа с целью ликвидации операции умножения при аппаратной или программной реализации банка фильтров.

3.5.2 Четырех канальные ПУБФ

Результаты синтеза общего 4-х канального ПУБФ (3.3) с длиной фильтра равной 8 показаны ниже. Коэффициенты решетчатой структуры банка фильтров приведены в табл. 3.1, а соответствующая АЧХ - на рис. 3.14. Для данного банка фильтров коэффициент эффективности кодирования $CG = 8.1227$ дБ и минимальное ослабление в полосе задержки составляет -20 дБ, что вполне согласуется с результатами из [21]. Из анализа АЧХ (рис. 3.14) следует, что полученный ПУБФ действительно имеет регулярность 1-го порядка: кроме фильтра нижних частот все фильтры имеют нулевое значение АЧХ на частоте DC . С другой стороны, на комплексной плос-

кости корней (рис. 3.15) видно, что фильтр нижних частот имеет один ноль в каждой точке на единичной окружности, соответствующей зеркально элайзинговой частоте.

Табл. 3.1 – Коэффициенты решетчатой структуры банка фильтров

Коэффициенты	Re(\cdot)	Im _i (\cdot)	Im _j (\cdot)	Im _k (\cdot)	Длина слова (бит)
r_0	-45/128	9/16	31/128	-5/8	8
s_0	-11/16	-1/2	1/16	7/16	5
s_1	3/8	1/8	3/4	-1/2	4

Другой пример расчета показывает эффективность предложенного подхода построения ПУБФ на кватернионах с линейной ФЧХ. В табл. 3.2 приведены коэффициенты банка фильтров (длина фильтра – 12), а на рис. 3.15 иллюстрируется АЧХ данного ПУБФ, из которого видно, что банк также имеет регулярность первого рода. Коэффициент эффективности кодирования и минимальное ослабление в полосе затухания соответственно равны: 8.1845 дБ и –20 дБ.

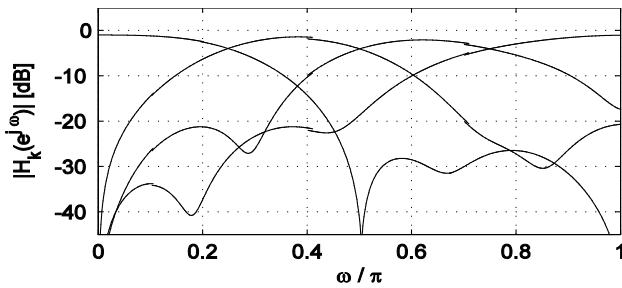


Рис. 3.14 – АЧХ 4-х канальный ПУБФ

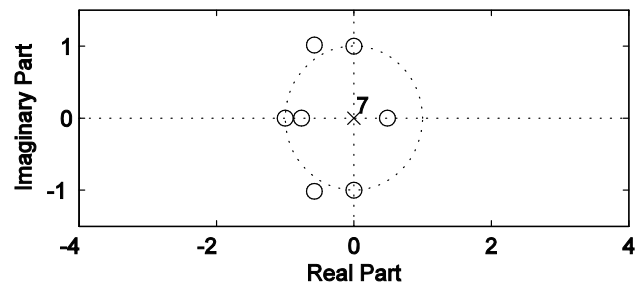


Рис. 3.15 – Плоскость корней ПУБФ

Табл. 3.2 – Коэффициенты банка фильтров с линейной ФЧХ

Коэффициенты	Re(\cdot)	Im _i (\cdot)	Im _j (\cdot)	Im _k (\cdot)	Длина слова (бит)
s_0	-231/512	459/1024	0	0	11
r_0	-7/8	-3/8	0	0	4
r_1	-3/16	15/16	0	0	5
r_2	-9/16	-13/16	0	0	5

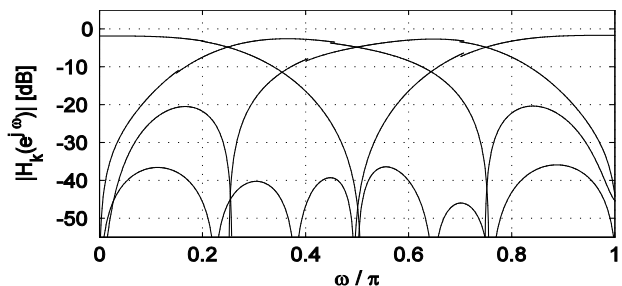


Рис. 3.16 – ПУБФ с линейной ФЧХ

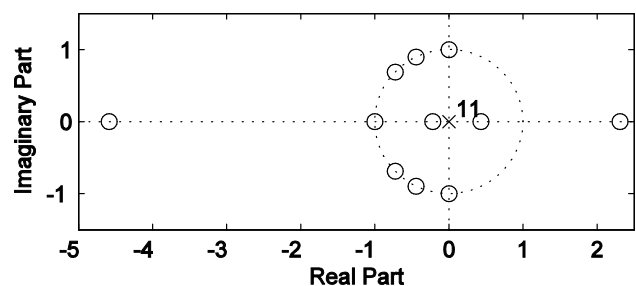


Рис. 3.17 – Плоскость корней ПУБФ

В табл. 3.3 и на рис. 3.18 и рис. 3.19 иллюстрируются результаты расчета 4-х канального ПУБФ с попарно зеркальной симметрией частотных характеристик фильтров на кватернионах (длина фильтра – 12), которые очень близки к результатам предыдущего примера: $CG =$

8.1699 дБ, минимальное ослабление в полосе затухания – 20 дБ. Это объясняется схожестью расположения нулей на плоскости корней (сравни рис. 3.17 и рис. 3.19)

Табл. 3.3 – Коэффициенты ПУБФ с попарно зеркальной симметрией частотных характеристик

Коэффициенты.	Re(\cdot)	Im _i (\cdot)	Im _j (\cdot)	Im _k (\cdot)	Длина слова (бит)
r_0	7/8	3/8	0	0	4
r_1	3/16	-1	0	0	5
r_2	-17/128	43/64	0	0	8

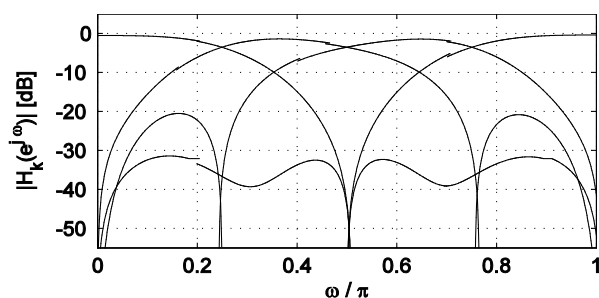


Рис. 3.18 – АЧХ ПУБФ с попарно зеркальной симметрией частотных характеристик фильтров

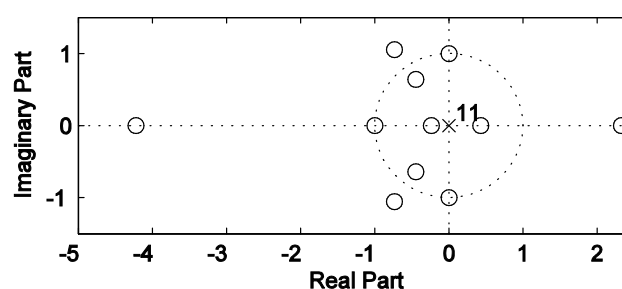


Рис. 3.19 – Плоскость корней ПУБФ

3.5.3 Восьми каналные ПУБФ

В данном параграфе рассматриваются результаты двух примеров расчета 8-канальных ПУБФ как для арифметики с неограниченной точностью, так и для арифметики с фиксированной запятой: с линейной ФЧХ и с попарно зеркальной симметрией частотных характеристик фильтров. В обоих случаях факторизация содержит три степени при длине фильтра 24.

Для первого банка фильтров с линейной ФЧХ и коэффициентами (см. табл. 3.4), представленными в арифметике с неограниченной точностью, коэффициент эффективности кодирования $CG = 9.3747$ дБ, а минимальное ослабление в полосе затухания равно – 21 дБ, что видно из рис. 3.20.

Табл. 3.4 – Коэффициенты 8-ми канального ПУБФ с линейной ФНЧ

Коэффициенты	Re(\cdot)	Im _i (\cdot)	Im _j (\cdot)	Im _k (\cdot)
s_0	-0.0297993	0.0669165	0.8901262	-0.449788
p_0	-0.2784728	-0.5164164	0.6659689	-0.4607087
q_0	-0.9687160	-0.0713704	-0.2281805	0.0665523
q_1	0.2571736	-0.2085346	-0.8265446	0.4551912
q_2	0.8835067	-0.2654561	0.0588609	-0.3814242
r_0	-0.1264641	0.6558758	-0.1001648	-0.7374285
r_1	0.9928326	0.0690029	-0.0687024	-0.0692971
r_2	0.9069558	0.3701135	-0.1523974	-0.1312336

В табл. 3.5 приведены коэффициенты второго банка фильтров, которые являются квантованными коэффициентами первого банка фильтров. Как видно из табл. 3.5, реализация данного ПУБФ не потребует умножителей. При этом, $CG = 9.3040$ дБ практически остался прежним, однако минимальное ослабление в полосе затухания уменьшилось до -13 дБ, как показано на рис. 3.22.

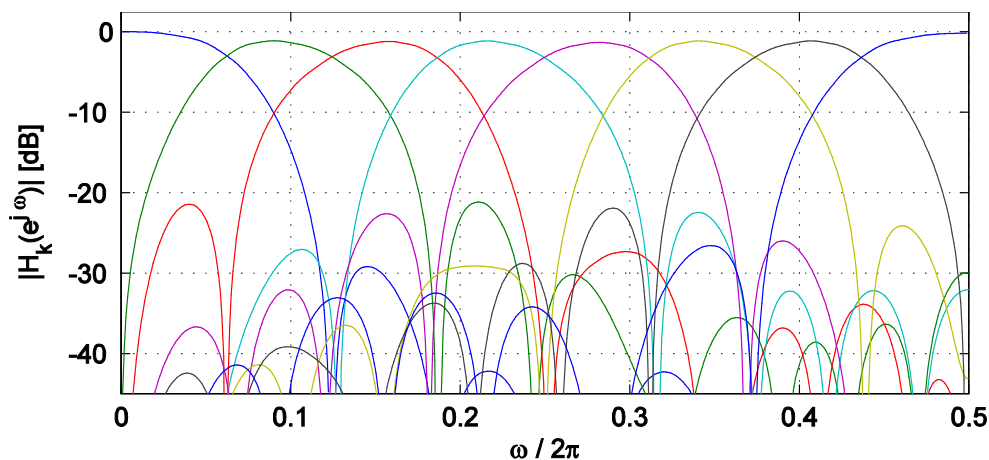


Рис. 3.20 – АЧХ 8-канальных ПУБФ с линейной ФЧХ

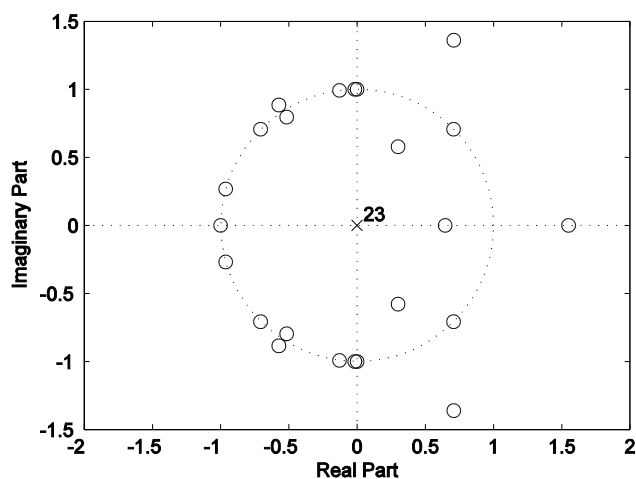


Рис. 3.21 – Плоскость корней ПУБФ

Табл. 3.5 – Квантованные коэффициенты 8-ми канального ПУБФ с линейной ФЧХ

Коэффициенты	$Re(\cdot)$	$Im_i(\cdot)$	$Im_j(\cdot)$	$Im_k(\cdot)$	Длина слова (бит)
s_0	$1/2$	0	$7/8$	$-1/8$	4
p_0	$-1/4$	$-1/2$	$5/8$	$-1/2$	4
q_0	$-15/16$	$-1/16$	$-1/4$	0	5
q_1	$1/4$	$-1/4$	$-7/8$	$1/2$	4
q_2	$7/8$	$-1/4$	$1/8$	$-3/8$	4
r_0	$-1/16$	$15/16$	$1/4$	$-1/4$	5
r_1	$15/16$	$1/16$	$-1/16$	$-1/16$	5
r_2	$3820719/2^{22}$	$611699/2^{21}$	$-54443/2^{19}$	$-435031/2^{22}$	23

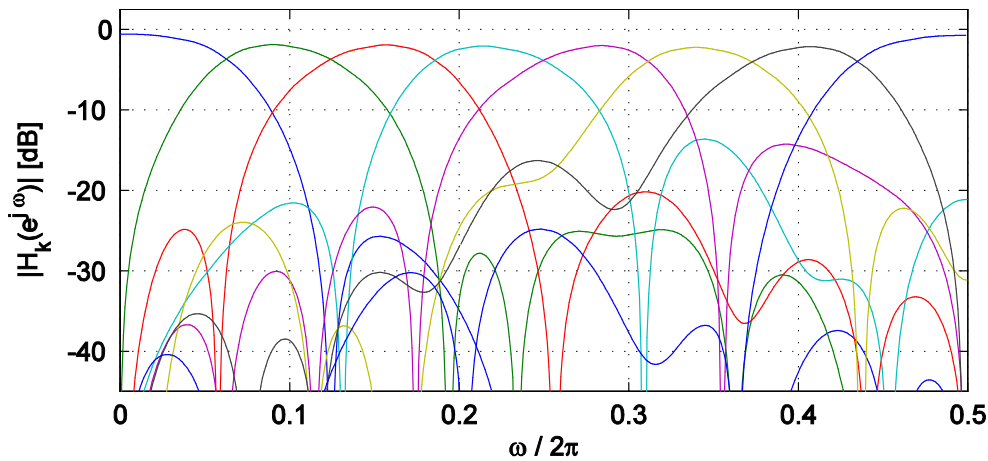


Рис. 3.22 – АЧХ 8-канальных ПУБФ с линейной ФЧХ с квантованными коэффициентами

Ниже приведены листинги программ в среде MATLAB вычисления импульсных и частотных характеристик 8-ми канального ПУБФ.

```

function h = hq8lp8(p0, q, r, s0);
% Computes the matrix of impulse responses of quaternionic structure of
% 8-channel linear-phase paraunitary filter bank
% for given hypercomplex coefficients.
% Both P0 and S0 are 4-element vectors (single quaternions)
% Q and R are 4 x K matrices (K quaternions, one column for each) where K
% is the number of stages.

%-----
% Auxiliary routines:

% QUATCNJ(q): conjugate of a given quaternion 'q' (4-element vector)
% QUATMULMTXL(q): left-multiplication matrix of a given quaternion 'q'
% QUATMULMTXR(q): right-multiplication matrix of a given quaternion 'q'

%-----
% Auxiliary matrices

I4 = eye(4);
J4 = fliplr(I4);
W8 = [I4 I4; I4 -I4];
W8_2 = W8 ./ 2;

%-----
% Initial stage

L = 8;
h = W8 * blkdiag(I4, J4) / sqrt(4);
h = blkdiag(quatmulmtxr(s0), quatmulmtxr(s0)) ...
    * blkdiag(quatmulmtxr(quadcnj(r(:, 1))), quatmulmtxr(r(:, 1))) ...
    * blkdiag(quatmulmtxl(quadcnj(q(:, 1))), quatmulmtxl(q(:, 1))) ...
    * blkdiag(quatmulmtxl(p0), quatmulmtxl(p0)) * h;

%-----

```

```

% Subsequent stages

for k = 2 : (size(q, 2)),
    L = L + 8;          % each stage extends filters by 8 taps

    h = W8 * h;

    t = zeros(8, L);   % matrix for extended impulse responses

    % delay: the first half of shorter responses advance
    % the second half by 8 taps

    t(1 : 4, 1 : (L - 8)) = h(1 : 4, :);
    t(5 : 8, 9 : L) = h(5 : 8, :);
    h = t;

    h = blkdiag(quatmulmtxr(quadconj(r(:, k))), quatmulmtxr(r(:, k))) ...
        * blkdiag(quatmulmtxl(quadconj(q(:, k))), quatmulmtxl(q(:, k))) ...
        * W8_2 * h;
end
h = h ./ 2;

clear all;

% preparing coefficients

s0 = [ 1/2, 0, 7/8, -1/8 ]';
p0 = [ -1/4, -1/2, 5/8, -1/2 ]';
q = [ -15/16, -1/16, -1/4, 0 ;
      1/4, -1/4, -7/8, 1/2 ;
      7/8, -1/4, 1/8, -3/8 ]';
r = [ -1/16, 15/16, 1/4, -1/4 ;
      15/16, 1/16, -1/16, -1/16 ;
      3820719/(2^22), 611699/(2^21), -54443/(2^19), -435031/(2^22) ]';

% computing impulse responses
h = hq8lp8(p0, q, r, s0);

% computing magnitude responses
N = 512;
H = fft(h, N, 2);
Hm_db = 20 * log10(abs(H));

% normalized frequency values
f = (0 : (N - 1)) / N;

% plotting magnitude responses
plot(f, Hm_db')
grid on
axis([0 0.5 -60 5])

function y = hyper_conj(x);
% conjugate of hypercomplex number

```

```

% x hypercomplex numbers (in columns)

y = x;
y(2 : end, :) = -x(2 : end, :);

function M = hyper_lmulptx(x);
% construction of left-multiplication matrix for hypercomplex number
% x hypercomplex number
% M multiplication matrix

M = [ x(1) -x(2) -x(3) -x(4); ...
      x(2) x(1) -x(4) x(3); ...
      x(3) x(4) x(1) -x(2); ...
      x(4) -x(3) x(2) x(1) ];

```

Следует отметить, оба банка являются банками с регулярностью первого порядка, потому что их кватернионные коэффициенты (см. табл. 3.4 и табл. 3.5) соответствуют условию (3.60). Анализ АЧХ (рис. 3.22 и рис. 3.20) показывает, что в данных ПУБФ не проявляется эффект *DC leakage* (см. п.3.2). На комплексной плоскости корней (рис. 3.21) видно, что фильтр нижних частот имеет один ноль в каждой точке на единичной окружности, соответствующей зеркально элайзинговым частотам.

Синтез второго банка фильтров с попарно зеркальной симметрией частотных характеристик фильтров приводится сначала в арифметике с неограниченной точностью (табл. 3.6): коэффициент эффективности кодирования $CG = 9.3655$ дБ, минимальное ослабление в полосе затухания равно -22 дБ (рис. 3.24), а затем коэффициенты данного банка представляются в арифметике с фиксированной запятой – в рациональных числах (табл. 3.7), что позволяет избежать умножителей при практической реализации банка.

Табл. 3.6 – Коэффициенты 8-ми канального ПУБФ с попарно зеркальной симметрией частотных характеристик фильтров

Коэффициенты	Re(\cdot)	Im _i (\cdot)	Im _j (\cdot)	Im _k (\cdot)
p_0	-0.0546530	-0.8381704	-0.2287760	0.4920823
p_1	0.9557711	0.0614376	0.1436822	0.2491636
p_2	-0.6224147	0.0345565	-0.5578179	-0.5479461
q_0	0.9916213	0.0606809	-0.0625301	-0.0953675
q_1	-0.0144669	0.6629532	0.1228110	0.7383774
q_2	-0.9805812	-0.0261340	0.0941266	0.1700519

Квантование кватернионных коэффициентов (табл. 3.7), в данном случае, привело к незначительному уменьшению коэффициента эффективности кодирования $CG = 9.2549$ дБ и значения минимального ослабления в полосе затухания -19 дБ (рис. 3.25). Учет условия (3.65) при синтезе банка фильтров с попарно зеркальной симметрией частотных характеристик фильтров гарантирует отсутствие искажений, обусловленных постоянной составляющей, а также анализ расположения нулей на комплексной плоскости корней показывает их корректность локализации (рис. 3.24).

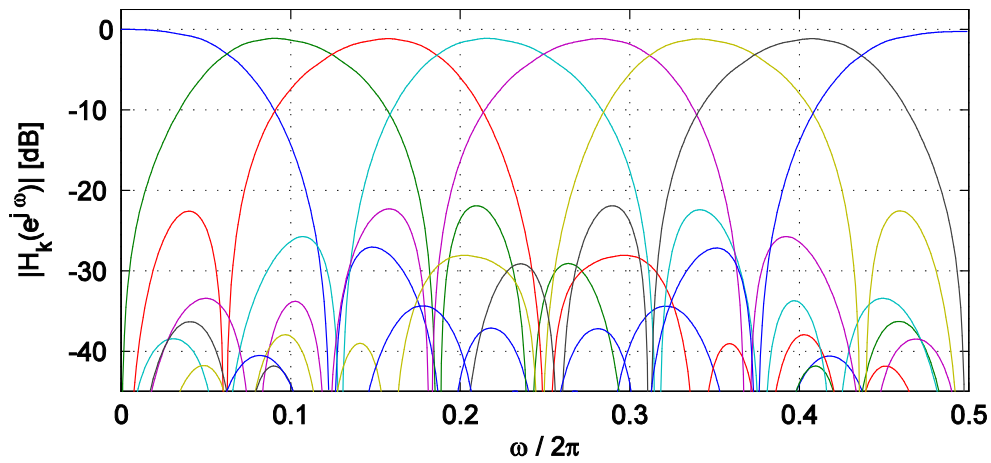


Рис. 3.23 – АЧХ 8-канальных ПУБФ с попарно зеркальной симметрией частотных характеристик фильтров

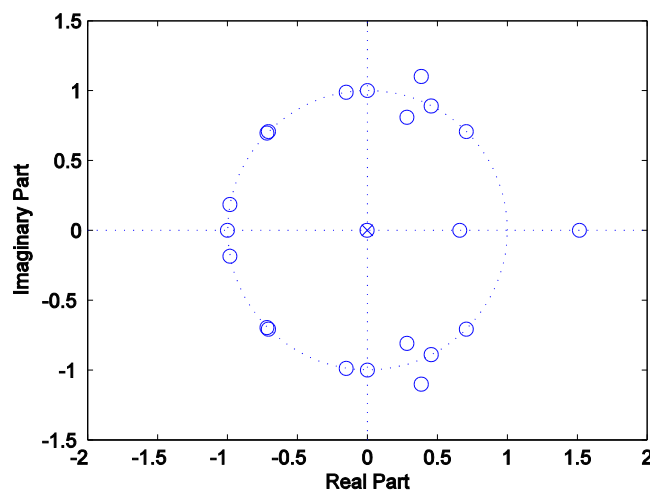


Рис. 3.24 – Комплексной плоскости корней ПУБФ

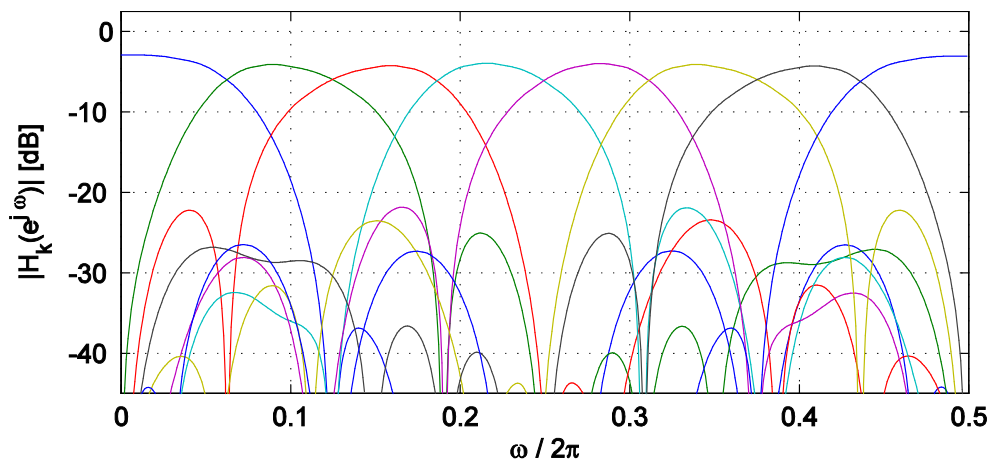


Рис. 3.25 – АЧХ 8-канальных ПУБФ с попарно зеркальной симметрией частотных характеристик фильтров

Табл. 3.7 – Квантованные коэффициенты 8-ми канального ПУБФ с попарно зеркальной симметрией частотных характеристик фильтров

Коэффициенты	$\text{Re}(\cdot)$	$\text{Im}_i(\cdot)$	$\text{Im}_j(\cdot)$	$\text{Im}_k(\cdot)$	Длина слова (бит)
p_0	$-1/8$	$-7/8$	$-1/4$	$1/2$	4
p_1	$7/8$	0	$1/8$	$1/4$	4
p_2	$-4307/2^{13}$	$1349/2^{14}$	$-8563/2^{14}$	$-3261/2^{13}$	15
q_0	$7/8$	0	0	$-1/8$	4
q_1	0	$5/8$	$1/8$	$3/4$	4
q_2	-1	$-1/16$	$1/16$	$1/8$	5

3.6 Заключение

В данной главе рассмотрен новый подход проектирования и реализации 4-х и 8-ми канальных ПУБФ на основе теории гиперкомплексных чисел, в частности, кватернионов, которая еще не применялась в области синтеза и анализа цифровых банков фильтров. Здесь умножитель кватернионов представлен как альтернативный модуль построения ПУБФ и может рассматриваться как обобщение планарных оборотов Гивенса. Представленные решетчатые структуры ПУБФ на кватернионах являются системами без потерь, несмотря на квантование коэффициентов и могут рассматриваться как расширение классических двух канальных решетчатых структур, разработанных Вайданасаном и Хоангом. Более того, предложенный подход дает возможность непосредственного выражения условия регулярности первого порядка, которое зафиксировано в значениях коэффициентов решетчатой структуры ПУБФ и, следовательно, может легко удовлетворяться даже для арифметики с конечной точностью вычислений. Таким образом, данная теория синтеза решетчатых структур ПУБФ на кватернионах представляет большой интерес с практической точки зрения обработки сигналов и является хорошей альтернативой стандартной технике построения ПУБФ [22,23].

3.7 Список использованных источников

1. Vaidyanathan P.P., Doganata Z., The role of lossless systems in modern digital signal processing: a tutorial, IEEE Trans. on Education, vol. 32, no. 3, pp. 181-197, 1989.
2. Vaidyanathan P.P., Multirate systems and filter banks, Prentice-Hall, Englewood Cliffs, NJ, USA, 1993.
3. Crochiere R.E., Rabiner L., Multirate digital signal processing, Prentice-Hall, Englewood Cliffs, NJ, USA, 1983.
4. Витязев В.В., Цифровая частотная селекция сигналов, Радио и связь, М., 1993, - 240с.
5. Piotrowski A., Parfieniuk M., Cyfrowe banki filtrow: analiza, synteza i implementacja dla systemow multimedialnych, Politechnika Bialostocka, Bialystok, 2006, - 389p.

-
6. Doganata Z., Vaidyanathan P.P., Nguyen T.Q., General synthesis procedures for FIR lossless transfer matrices for perfect-reconstruction multirate filter bank applications, *IEEE Trans. On Acoustics, speech and signal processing*, vol. 36, n0. 10, pp. 1561-1574, 1988.
 7. Gan L., Ma K. –K., A simplified lattice factorization for linear-phase perfect reconstruction filter bank, *IEEE Signal processing letters*, vol. 8, n0. 7, pp. 207-209, 2001.
 8. Nguyen T.Q., Vaidyanathan P.P., Maximally decimated perfect-reconstruction FIR filter banks with pairwise mirror-image analysis (and synthesis) frequency responses, *IEEE Trans. On Acoustics, speech and signal processing*, vol. 36, no. 5, pp. 693-706, 1988.
 9. Oraintara S., Tran T.D., Heller P.N., Nguyen T.Q., Lattice structure for regular paraunitary linear-phase filterbanks and M-band orthogonal symmetric wavelets, *IEEE Trans. On Signal processing*, vol. 49, no. 11, pp. 2659-2672, 2001.
 10. Vaidyanathan P.P., On coefficient-quantization and computational roundoff effects in lossless multirate filter banks, *IEEE Trans. On Signal processing*, vol. 39, no. 4, pp. 1006-1008, 1991.
 11. Hamilton W.R., On quaternions; or on a new system of imaginaries in algebra, *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, vol. 25, pp.489-495, 1844.
 12. Kantor I.L., Solodovnikov A.S., *Hypercomplex numbers: an elementary introduction to algebra*, Springer, New York, NY, USA, 1989.
 13. Baker A., *Matrix groups: an introduction to Lie group theory*, Springer, London, UK, 2002.
 14. Baker H.G., Quaternions and orthogonal 4x4 real matrices, Tech. Rep., June 1996, <http://www.gamedev.net/reference/articles/article428.asp>.
 15. Parfieniuk M., Petrovsky A., Hypercomplex factorizations for 8-channel linear phase paraunitary filter banks, in Proc. 7th Int. Conf. and exhibition “Digital signal processing and its applications” (DSPA), vol. 2, Moscow, Russia, 2005, pp. 509-513.
 16. Parfieniuk M., Petrovsky A., Quaternionic approach to 8-channel general paraunitary filter banks, in Proc. 13th European Signal processing conference (EUSIPCO), Antalya, Turkey, 2005, CD.
 17. Howell T.D., Lafon J.C., The complexity of the quaternion product, Tech. Rep. TR 75-245, Cornell University, Ithaca, NY, USA, June 1975, <http://citeseer.ist.psu.edu/howell75complexity.html>.
 18. Verenik A., Parfieniuk M., Petrovsky A., An FPGA implementation of the distributed arithmetic based quaternionic multipliers for paraunitary filter banks, in Proc. of the MIXDES’2007, Ciechocinek, Poland, 2001, pp. 605-610.
 19. Parfieniuk M., Petrovsky A., Implementation perspectives of quaternionic component for paraunitary filter banks, in Proc. of the International Workshop on Spectral methods and multi-rate signal processing (SMMSP’04), pp. 151-158, Vienna, Austria, Sep. 2004.

-
20. Parfieniuk M., Petrovsky A., Quaternionic lattice structures for four-channel paraunitary filter banks, EURASIP Journal on Advances in Signal Processing, vol 2007, Article ID 37481, 12pages, 2007.
 21. Chen Y.-J., Oraintara S., Amaratunga K.S., Dyadic-based factorizations for regular paraunitary filterbanks and M-band orthogonal wavelets with structural vanishing moments, IEEE Trans. on Signal processing, vol. 53, no. 1, pp. 193-207, 2005.
 22. Парфенюк М., Петровский А.А., Параунитарные банки фильтров на основе алгебры кватернионов: теория и применение, Научно-технический журнал «Цифровая обработка сигналов», № 1, с.22-36, 2008, №2, с.57, 2008.
 23. Parfieniuk M., Petrovsky A., Inherently lossless structures for eight- and six- channel linear-phase paraunitary filter banks based on quaternion multiplier, Signal processing, vol. 90, no. 6, pp. 1755-1767.

4 АНАЛИЗ ЗВУКОВЫХ СИГНАЛОВ С АДАПТИВНЫМ ЧАСТОТНО-ВРЕМЕННЫМ ПЛАНОМ

Петровский Ал.А.

АННОТАЦИЯ

Теоретическое направление цифровой обработки сигналов - дискретное вейвлет преобразование, в отличие от известных, позволяет осуществлять обработку нестационарных сигналов в частотно-временной области, проводить анализ с переменным частотным разрешением сигналов. На базе вейвлет преобразования возможно построение пакета адаптивного вейвлет преобразования, с помощью которого осуществляется эффективная обработка звуковых сигналов. В данной статье описывается вычисление дискретного вейвлет преобразования, построение пакета вейвлет преобразования, а также показаны алгоритмы адаптации декомпозиции дискретного вейвлет преобразования в входному сигналу. Задачей данной статьи является введение в дискретное вейвлет преобразование и демонстрация особенностей и преимуществ нового преобразования нестационарных сигналов на основе алгоритма согласованной подгонки.

Ключевые слова – вейвлет преобразование, адаптивный пакет дискретного вейвлет преобразования, алгоритм согласованной подгонки.

СОДЕРЖАНИЕ

4.1	Вейвлет преобразование	74
4.1.1	Многомаштабный анализ	74
4.1.2	Вейвлет преобразование и пакет дискретного вейвлет преобразования (ПДВП)	76
4.1.3	Вейвлет преобразование с переменным временным шагом	82
4.2	Методы декомпозиция ПДВП	84
4.2.1	Процедура роста дерева ПДВП.....	84
4.2.2	Выбор фильтров прототипов для ПДВП.....	86
4.2.3	Декомпозиция ПДВП на основе алгоритма «простого дерева»	90
4.2.4	Декомпозиция ПДВП на основе алгоритма «двойного дерева».....	91
4.2.5	Декомпозиция ПДВП на основе алгоритма «частотно-временного дерева».....	92
4.2.6	Декомпозиция ПДВП на основе перцептуальной энтропии.....	93
4.3	Согласованная подгонка на основе адаптивного ПДВП	94
4.4	Список использованных источников.....	96

4.1 Вейвлет преобразование

4.1.1 Многомаштабный анализ

Звуковой сигнал $x(t) \in L^2(k)$ при частотно-временном анализе с переменной разрешающей способностью как по частоте (масштабе) так и по времени, другими словами много-

масштабный анализ [1-3] представляется нарастающей последовательностью $V_{l,n}$ при $n = 0$, $V_{l,n} = V_l$, $l, n \in Z$ замкнутых линейных подпространств $L^2(k)$, т.е.

$$\dots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \dots \quad (4.1)$$

Эти пространства имеют следующие свойства:

- полноты $\bigcap_{l,n \in Z} V_{l,n} = \{0\}$, $\bigcup_{l,n \in Z} V_{l,n}$ всюду плотно в $L^2(k)$;
- масштабируемости: для всех сигналов $x(t) \in L^2(k)$ и всех $l, n \in Z$ их сжатая версия будет принадлежать пространству V_{l-1}

$$x(t) \in V_l \leftrightarrow x(2t) \in V_{l+1}, \forall x(t) \in L^2(k), \quad (4.2)$$

- сдвига: для всех $x(t) \in L^2(k)$ и всех $k \in Z$

$$x(t) \in V_0 \leftrightarrow x(t - k) \in V_0, \quad (4.3)$$

- ортонормальности: существует такая функция $\varphi_{0,n,k}(t) \in V_0$, что её сдвиги $\varphi_{0,n,k}(t) = \varphi_n(t - k)$, $n, k \in Z$ образуют ортонормальный базис пространства V_0 . Так как функции $\varphi_{0,n,k}(t)$ образуют ортонормальный базис пространства V_0 , то функции

$$\varphi_{l,n,k}(t) = 2^{-\frac{l}{2}} \varphi_n(2^{-l}t - k), \forall l \in Z, \quad (4.4)$$

образуют ортонормальный базис пространства V_l . Эти базисные функции называются масштабными функциями, так как они создают масштабные версии функций в $L^2(k)$.

Таким образом, многомасштабный анализ проводится с помощью набора последовательных аппроксимирующих пространств V_l , которые представляют собой отмасштабированные и инвариантные относительно смещения на целые числа разновидности одного центрального функционального пространства V_0 . В этом пространстве существует ортонормальный базис. Так сигнал $x(t)$ в $L^2(k)$ может быть представлен множеством последовательных его преобразований $x_l(t)$ в V_l [4]. Другими словами, сигнал $x(t)$ есть предел аппроксимации $x_l(t) \in V_l$ при $l \rightarrow -\infty$:

$$x(t) = \lim_{l \rightarrow -\infty} x_l(t). \quad (4.5)$$

Отсюда появляется возможность анализа сигнала на различных уровнях частотного разрешения или масштаба. Переменная l называется масштабным коэффициентом или уровнем анализа. Если значение l мало, то функция V_l есть грубая аппроксимация $x(t)$ и детали отсутствуют. При больших значениях l имеет место точная аппроксимация.

Функция $\varphi_{l,n,k}(t)$ образует ортонормальный базис V_l . Ортонормальное дополнение V_l в V_{l-1} обозначается $V_{l,n}$ при $0 < n < 2^l$. Подпространства $V_{l,n}$ образуют взаимно ортогональный набор. Последовательность $\psi_{l,n,k}(t)$ формирует ортонормальный базис $V_{l,n}$ при любом заданном l . Полный набор функций $\varphi_{l,n,k}(t)$ и $\psi_{l,n,k}(t)$ при всех l образуют ортонормальный базис $L^2(k)$. На рис. 4.1 показано многомасштабное пространство $L^2(k)$

Масштабная функция $\varphi(t)$ ассоциируется с подпространством V_3 . Другая функция, которая ассоциируется с подпространством $V_{3,1}$, называется вейвлет функцией $\psi(t)$.

Функция $\psi_{l,n,k}(t)$ будет вейвлетом класса l , если выполняются следующие четыре свойства [5]:

- если $l > 0$, то $\psi_{l,n,k}(t)$ и $\varphi_{l,n,k}(t)$ принадлежат к пространству $L^2(k)$;
- если $l \geq 0$, то $\psi_{l,n,k}(t)$ и $\varphi_{l,n,k}(t)$, и все их производные вплоть до порядка l принадлежат к $L^2(k)$;
- $\psi_{l,n,k}(t)$, $\varphi_{l,n,k}(t)$ и все их производные вплоть до порядка l , быстро уменьшаются, когда $t \rightarrow \pm\infty$;

$$\int_{-\infty}^{+\infty} \psi_{l,n,k}(t) \partial t = 0, \text{ и } \int_{-\infty}^{+\infty} \varphi_{l,n,k}(t) \partial t = 0, \quad (4.6)$$

- набор функций $2^{l/2}\psi_n(2^l t - k)$, $2^{l/2}\varphi_n(2^l t - k)$, $l, n, k \in Z$, является ортонормальным базисом пространства $L^2(k)$.

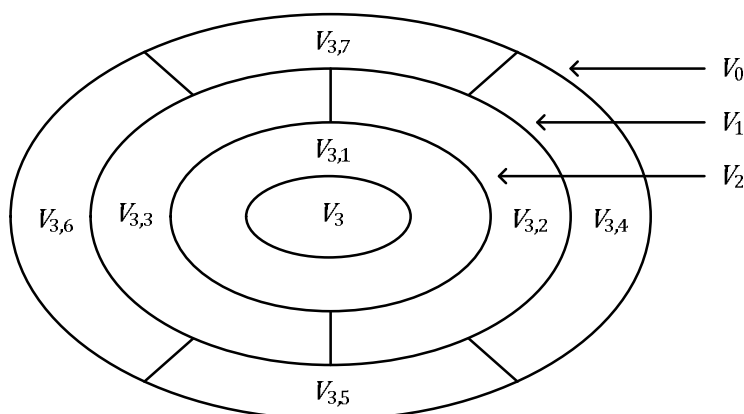


Рис. 4.1 – Многомасштабное пространство $L^2(k)$

Функции $2^{l/2}\psi_n(2^l t - k)$ являются вейвлетами, порожденными одним «материнским вейвлетом» (основным вейвлетом) ψ . Условия 1), 2) и 3) ответственны за регулярность, локальность и знакопеременность вейвлет функции $\psi(t)$.

Масштабная и вейвлет функции представляются следующим образом:

$$\varphi_{l,n,k}(t) = \sqrt{2} \sum_{k \in Z} h_0 \varphi_{l,n}(2t - k), \quad (4.7)$$

$$\psi_{l,n,k}(t) = \sqrt{2} \sum_{k \in Z} h_1 \psi_{l,n}(2t - k), \quad (4.8)$$

где $h_1(k) = \langle \psi_{l,n,k}(t), \varphi_{l-1,n,k}(t) \rangle$ и $h_0(k) = \langle \varphi_{l,n,k}(t), \varphi_{l-1,n,k}(t) \rangle$ - коэффициенты соответственно высокочастотного и низкочастотного фильтров. При цифровой реализации это цифровые КИХ-фильтры.

4.1.2 Вейвлет преобразование и пакет дискретного вейвлет преобразования (ПДВП)

Из определения многомасштабного анализа следует, что любой сигнал $x(t) \in L^2(k)$ может быть представлен как линейная комбинация масштабных и вейвлет функций, т.е. как сумма его ортогональных проекций на подпространства $V_{l,n}$ [2]:

$$x(t) = \sum_{k \in \mathbb{Z}} c_{l,n}(k) \cdot \varphi_{l,n,k}(t) + \sum_{l \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} d_{l,n}(k) \cdot \psi_{l,n,k}(t), \quad (4.9)$$

где

$$c_{l,n}(k) = \sum_{m \in \mathbb{Z}} h_0(m - 2k) c_{l-1,n}(k), \quad (4.10)$$

и

$$d_{l,n}(k) = \sum_{m \in \mathbb{Z}} h_1(m - 2k) c_{l-1,n}(k). \quad (4.11)$$

Если $c_{l-1,n}(k)$ представляют собой отсчеты дискретного сигнала, то итерационная процедура (4.10) декомпозиции $c_{l,n}(k)$ определяет дискретное вейвлет преобразование, которое на этапе анализа интерпретируется как банк фильтров-дециматоров, а на этапе синтеза как банк фильтров-интерполяторов (рис. 4.2):

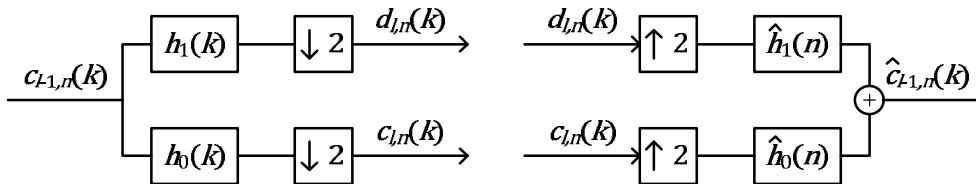
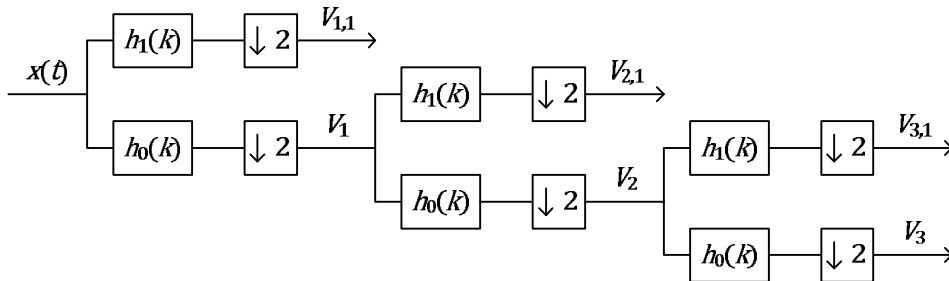
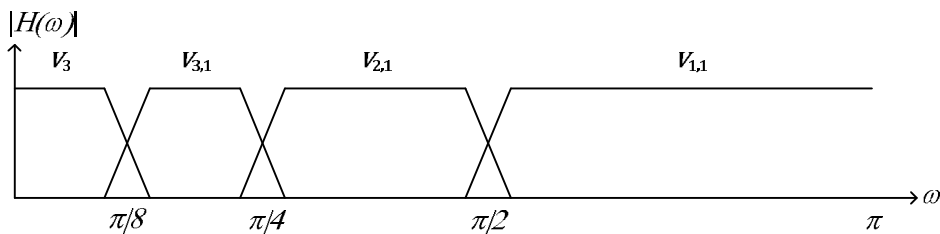


Рис. 4.2 – Банки фильтров-дециматоров и фильтров-интерполяторов



а)



б)

Рис. 4.3 – Трехступенчатое дерево анализа а) и соответствующее ему разбиение на частотные полосы б)

Итак, схема (рис. 4.2) делит сигнал уровня $l = 0$ на два сигнала уровня $l = 1$. Далее вейвлет преобразование получается путем рекурсивного применения данной схемы к низко-

частотной части. Например, логарифмическое распределение полос, т.е. фильтрация с постоянной добротностью Q (рис. 4.3).

Здесь вейвлет $2^{l/2}\psi_n(2^l t - k)$ «сконцентрирован» в основном на двоичном интервале

$$[n2^{-l}, (n + 1)2^{-l}] \text{ при } l, n \in Z. \quad (4.12)$$

Его Фурье-образ заключён в интервале

$$2^{-l}\pi n \leq |\omega| \leq 2^{-l}\pi(n + 1). \quad (4.13)$$

В действительности его частотная полоса покрывает одну октаву.

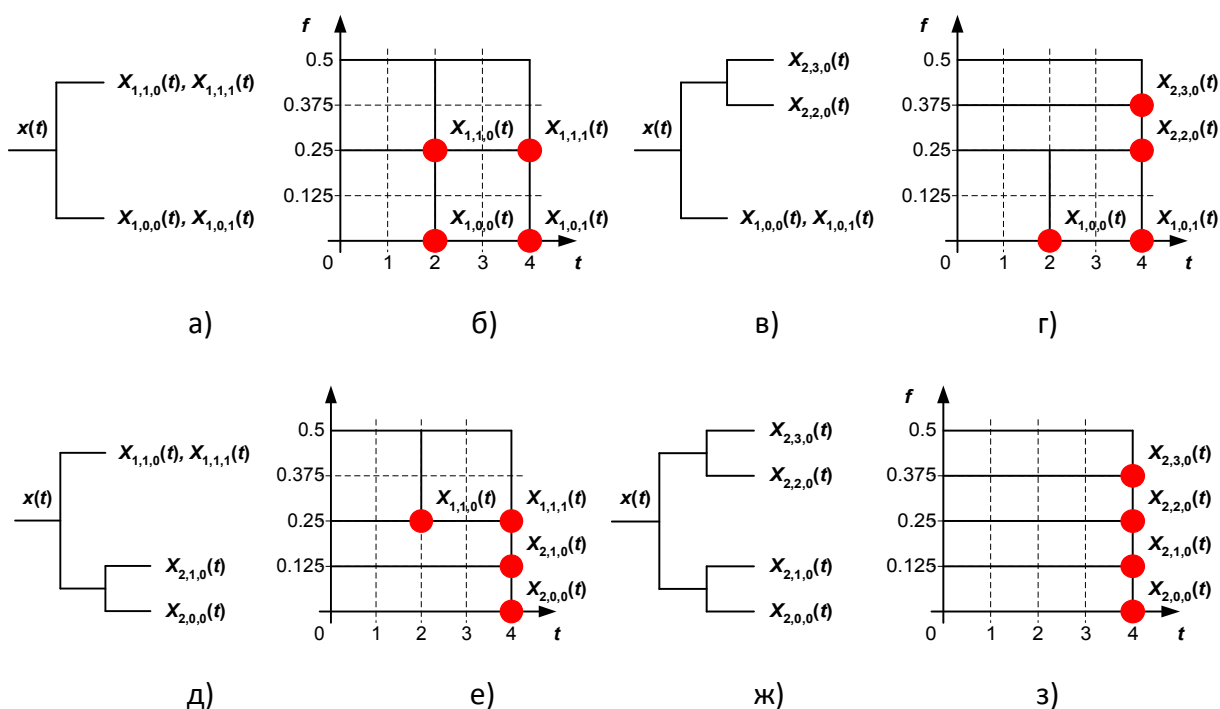


Рис. 4.4 – Декомпозиция сигнала на базе ПДВП

Вейвлет преобразование является сигналонезависимым. Октавополосное разбиение спектра, производимое им, подходит для большинства приложений, но не для всех реальных сигналов, в частности, звука. Желательно было бы иметь преобразование, адаптированное к сигналу. Это эквивалентно тому, что преобразование было бы способно произвольно менять структуру разбиения частотно-временной плоскости в зависимости от сигнала. Каскадно-соединенные блоки вейвлет фильтров позволяют достичь этого [6].

В отличие от дискретного вейвлет преобразования, представляющего собой дерево декомпозиции области нижних частот, ПДВП это декомпозиция обеих частей дерева преобразования: области нижних и верхних частот [2]. На рис. 4.4(а, в, д, ж) показаны возможные варианты построения второй ступени ПДВП.

Как видно из рис. 4.4 структура дерева преобразования позволяет варьировать разрешающую способность в частотной и временной областях. Выполнение более детальной декомпозиции, например, в области нижних частот, приводит к увеличению разрешающей способности в области частот и ее уменьшению во временной области, в то время как менее детальная декомпозиция в области верхних частот, обеспечивает высокое разрешение во временной об-

ласти и уменьшает частотное разрешение. Соответствующие частотно-временные планы показаны на рис. 4.4(б, г, е, з), где кружками отмечены коэффициенты преобразования. Здесь индексы имеют следующее назначение: i – уровень преобразования, j – номер узла на уровне i ; t – это временная позиция в блоке обработки.

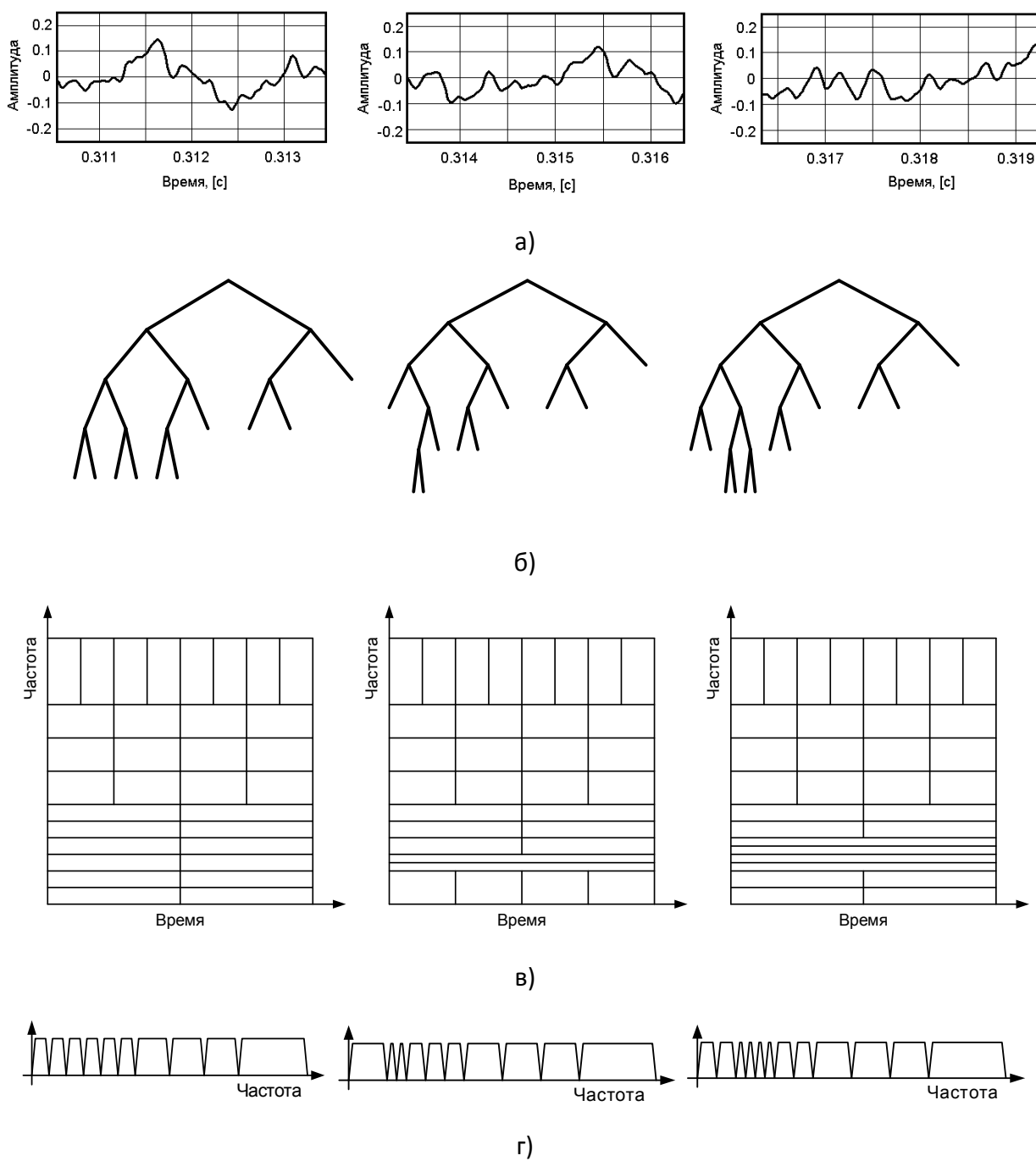


Рис. 4.5 – Пример структур деревьев ПДВП для трёх фреймов звукового сигнала

Далее на основе введенной функции стоимости определяется наилучший путь по этому дереву. Если исходный блок вейвлет фильтров был ортогональным, то и схема, соответствующая любой конфигурации, будет ортогональной, так как она есть не что иное, как каскадное соединение ортогональных блоков. Таким образом, получается базис, адаптированный к сигналу. Отметим, что эта адаптация не требует обучения или знания статистических свойств сигнала.

Вейвлет преобразование (рис. 4.4(ж, з)), является частным случаем этого базиса. Конечно, адаптивность достигается за счет увеличения вычислительной сложности алгоритмов [6].

На рис. 4.5 для разных фреймов звукового сигнала (а) показаны наиболее адекватные процессу обработки структуры деревьев ПДВП (б), а также их частотно-временные планы (в) и итоговое разбиение частотного диапазона на частотные полосы (г) в соответствии с локализацией энергии в исходном сигнале.

Таким образом, в отличие от традиционно применяемого для анализа сигналов преобразования Фурье и банка ДПФ фильтров ПДВП обеспечивает двумерную развертку кодируемого звукового сигнала, при этом частота и время рассматриваются как независимые переменные. В результате появляется возможность анализировать свойства сигнала одновременно во временном и частотном пространствах.

Программа вычисления декомпозиции сигнала на основе ПДВП.

```
function [data, tree] = wpdec(x,tree)
%WPDEC Wavelet packet decomposition 1-D.
%      [D,S] = WPDEC(X,N,'wname') returns a data structure D,
%      corresponding to a
%      wavelet packet decomposition of the vector X, at levels N,
%      and a sizes of each part of decomposition
%      with a particular wavelet ('wname', see WAVEFILTERS).

% Initialization
%-----
order      = 2;
depth      = fix(log10(max(tree.N))/log10(order));
tree.sizes = zeros(1,length(tree.N));
tree.length = length(x);
data       = x(:)';
un_incl    = tree.N;
lx         = length(x);

% Tree computation
%-----
[tree.LoF_D,tree.HiF_D] = wavefilters(tree.wname,'d');
for k=0:depth-1
    beg = 1;
    for p=0:order^k-1
        if all(un_incl - (order^k+p) ~= 0,
            x = data(beg:beg+lx-1);
            [a,d] = dwt(x,tree.LoF_D,tree.HiF_D);
            data = [data(1:beg-1) a d data(beg+lx:length(data))];
            beg = beg+2*length(a);
            if k == depth-1,
                tree.sizes = [tree.sizes(1:find(tree.N == ...
                    (order^k+p)*order)-1) length(a) length(d) ...
                    tree.sizes(find(tree.N == ...
                    (order^k+p)*order)+2:length(tree.sizes))];
            end;
        else
```

```

    beg = beg+lx;
    un_incl = [un_incl(1:find(un_incl == (order^k+p))-1) ...
              order^(k+1)+order*(p) order^(k+1)+order*(p)+1 ...
              un_incl(find(un_incl == (order^k+p))+1:length(un_incl))];
    if max(tree.N == (order^k+p)) ~= 0,
        tree.sizes = [tree.sizes(1:find(tree.N == ...
                                      (order^k+p))-1) lx tree.sizes(find(tree.N ...
                                      == (order^k+p))+1:length(tree.sizes))];
    end;
end;
end
lx = length(a);
end

```

Программа вычисления реконструкции ПДВП.

```

function [y] = wprec(data, tree)
%WPREC Wavelet packet reconstruction 1-D.
%      X = WPREC(T,D) returns the reconstructed vector
%      corresponding to a wavelet packet decomposition
%      structure [T,D] (see MAKETREE).
%
%      T is the tree structure and D the data structure.

% Initialization
%-----
order      = 2;
depth     = fix(log10(max(tree.N))/log10(order));

% Tree computation
%-----
[LoF_R,HiF_R] = wavefilters(tree.wname,'r');
for k=depth:-1:1,
    beg = 1;
    ly = 2*min(tree.sizes)-length(LoF_R);
    for p=0:2:order^k-1,
        if max(tree.N == (order^k+p)) ~= 0,
            a = data(beg:beg+tree.sizes(find(tree.N == (order^k+p))-1));
            d = data(beg+tree.sizes(find(tree.N == ...
                                      (order^k+p))):beg+tree.sizes(find(tree.N == ...
                                      (order^k+p))+tree.sizes(find(tree.N == (order^k+p))+1)-1));
            if abs(tree.sizes(find(tree.N == (order^k+p)))-...
                tree.sizes(find(tree.N == (order^k+p))+1)) > 0,
                a = a(1:min([tree.sizes(find(tree.N==(order^k+p)) ...
                              tree.sizes(find(tree.N == (order^k+p))+1)]));
                d = d(1:min([tree.sizes(find(tree.N==(order^k+p)) ...
                              tree.sizes(find(tree.N == (order^k+p))+1)]));
            end;
            y = idwt(a,d,LoF_R,HiF_R);
            data = [data(1:beg-1) y data(beg+tree.sizes(find(tree.N==...
                                      (order^k+p))+tree.sizes(find(tree.N ==...
                                      (order^k+p))+1):length(data))];
        end;
    end;
end;

```

```

        beg = beg+length(y);
        tree.sizes = [tree.sizes(1:find(tree.N == (order^k+p))-1)...
        length(y) tree.sizes(find(tree.N ==...
        (order^k+p))+2:length(tree.sizes))];
        tree.N = [tree.N(1:find(tree.N == (order^k+p))-1)...
        fix(tree.N(find(tree.N == (order^k+p))/2)...
        tree.N(find(tree.N == (order^k+p))+2:length(tree.N))];
    else
        beg = beg + ly;
    end;
end;
end;
y = data(1:tree.length)';

```

4.1.3 Вейвлет преобразование с переменным временным шагом

Анализ входного сигнала $x(t)$ на базе пакета дискретного вейвлет преобразования (ПДВП) с переменным частотно-временным разрешением (многомасштабный анализ [1,2]) выполняется на основе рекурсивных функций (4.7) и (4.8), где $h_1(k) = (-1)^k h_0(1-k)$ и $\psi_{l,2n+1}(t) \equiv \varphi_{l,2n}(t)$ ортонормальные масштабные функции преобразования. При цифровой реализации это коэффициенты цифровых фильтров с конечной импульсной характеристикой. Декомпозиции структуры дерева ПДВП $l, n \in Z$ определяют множество всех возможных ортонормальных базисов спроецированных на подпространство $V_{l,n}$

$$V_{l,n} \equiv \left\{ \psi_{l,2n+1}(t) = 2^{l/2} \sum_{k \in Z} \varphi_{l,n}(2t - k) \right\} : l, n \in Z, \quad (4.14)$$

где l определяет текущий уровень декомпозиции, а n номер узла на уровне l . Из многомасштабного анализа следует, что любой сигнал $x(t) \in L^2(k)$ может быть представлен как линейная комбинация масштабных функций, т.е. как сумма его ортогональных проекций на подпространстве $V_{l,n}$ [2].

Декомпозиция ПДВП может быть расширена в корень из двух раз путем варьирования временным сдвигом в каждом из узлов n структуры дерева ПДВП $l, n \in E$. Увеличение числа возможных комбинаций не изменяет древовидную структуру ПДВП. Временной сдвиг модифицирует ортогональные проекции ортонормальных масштабных функций на подпространстве $V_{l,n}$ [1,7]. Данное изменение описывает как уже существующие, так и вновь образованные подмножества

$$V_{l,n,m} \equiv \left\{ \psi_{l,2n+1,m}(t) = 2^{l/2} \sum_{k \in Z} \varphi_{l,n,m}(2(t - m_c) - k) \right\} : l, n \in E, \quad (4.15)$$

где $m_c = \{m, m + 2^{-l}\}$ – временной сдвиг. Тройные индексы каждого узла (l, n, m) структуры дерева E ПДВП определяют номер узла n , уровень l , временной сдвиг m . На рис. 4.6 показаны декомпозиции вейвлет преобразования без временного сдвига и с ним, как это показано на рис. 4.7. ПДВП получается путем каскадного соединения блоков декомпозиций данных схем.

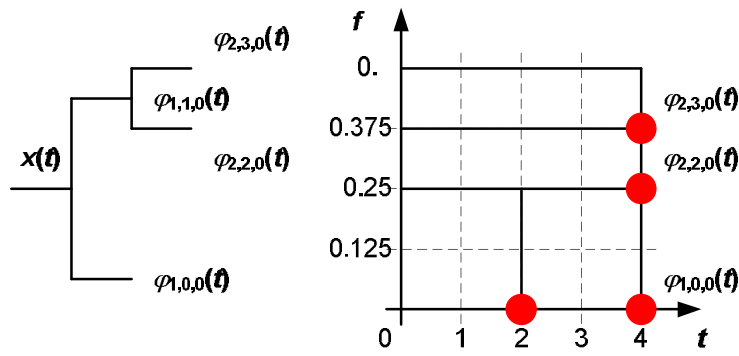


Рис. 4.6 – Декомпозиция ПДВП без временного сдвига

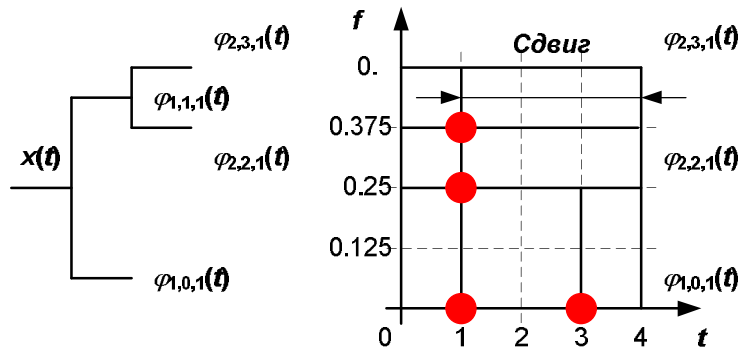


Рис. 4.7 – Декомпозиция ПДВП с временным сдвигом

Как видно из рис. 4.7 структура дерева ПДВП позволяет не только варьировать разрешающую способность в частотной и временной областях, но и выполнять сдвиг по времени, для более детальной декомпозиции входного сигнала [8].

Результаты моделирования показаны на рис. 4.8 и рис. 4.9. Тестовый сигнал, содержащий короткие импульсы тона и нелинейной частотной модуляции, анализировался на основе адаптивного ПДВП без временного сдвига (рис. 4.8) и с временным сдвигом (рис. 4.9).

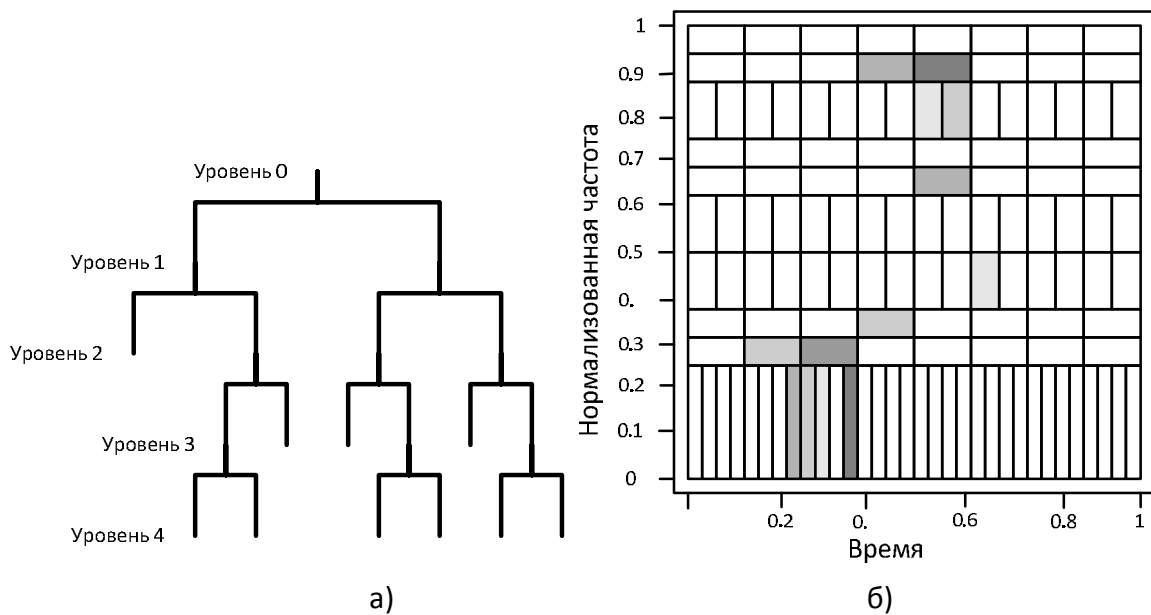


Рис. 4.8 – Структура ПДВП (а) и частотно-временной план ПДВП без временного сдвига (б)

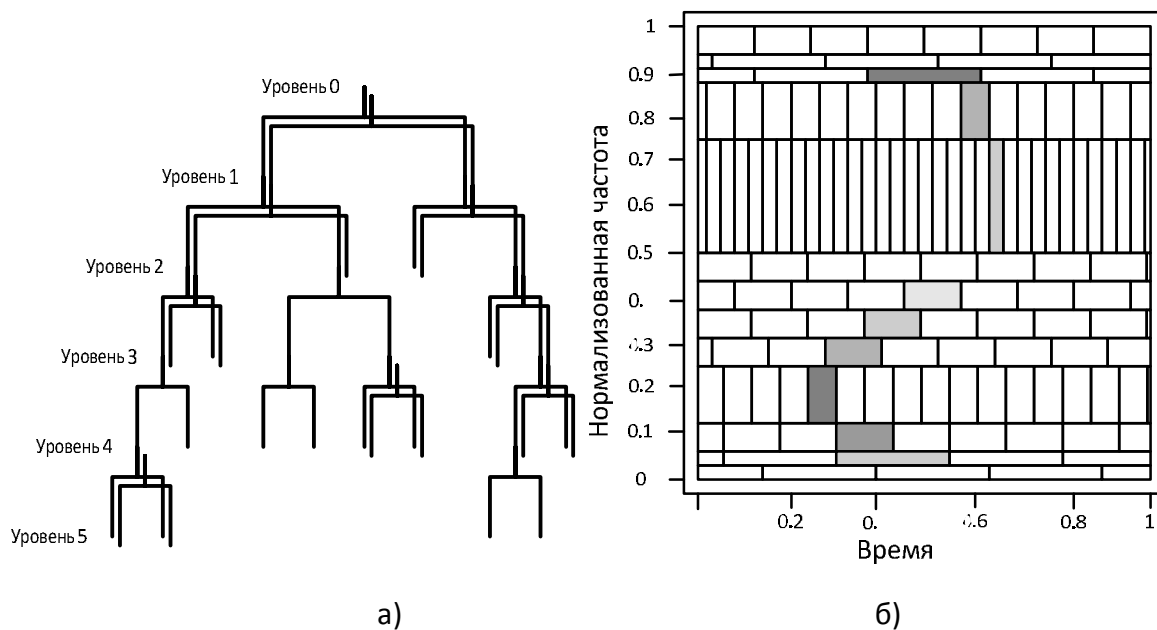


Рис. 4.9 – Структура ПДВП (а) и частотно-временной план ПДВП с временным сдвигом (б)

Результат, полученный на основе адаптивного ПДВП с временным сдвигом, лучше локализовал неопределенность (информативность) сигнала, чем без сдвига.

4.2 Методы декомпозиция ПДВП

4.2.1 Процедура роста дерева ПДВП

Поиск структур деревьев ПДВП базируется на утверждении, которое доказывает существование множества структур ПДВП. Следовательно, стоит задача поиска такой структуры дерева преобразования из библиотеки, при которой обеспечивается максимальная компрессия без воспринимаемых на слух вносимых искажений в реконструированный сигнал при заданном временном разрешении.

Структура банка фильтров (дерева преобразования ПДВП) определяет вычислительную сложность алгоритма процедур анализа и синтеза на каждом фрейме. При этом, необходимо вести оптимизацию дерева преобразования ещё и под заданный вычислительный ресурс.

Конструкция алгоритма выбора лучшей декомпозиции E ПДВП следующая [6,9,10]. Для каждого узла $(l, n) \in E$ дерева ПДВП вычисляется стоимостная функция J . Далее, рост структуры дерева осуществляется в соответствии со следующей процедурой:

Процедура роста структуры дерева ПДВП

Пусть $J_{\text{предок}}$ - мера качества узла предка дерева ПДВП,
 $J_{\text{потомки}}$ - сумма мер качества соответствующих узлов потомков,

тогда

если $J_{\text{предок}} > J_{\text{потомки}}$,
то $J_{\text{предок}} = J_{\text{потомки}}$,

иначе отсечь узлы потомков.

Коэффициенты $c_{l,n}(k)$ и $d_{l,n}(k)$ дерева E преобразования ПДВП на выходе из процедуры отсечения ветвей дерева будут соответствовать лучшей структуре ПДВП в смысле сформулированном выше.

Таким образом, ключевыми моментами алгоритма декомпозиции структуры дерева ПДВП являются:

- структура банка фильтров ПДВП, определяющая вычислительную сложность алгоритма преобразования на каждом масштабном уровне;
- алгоритм адаптации структуры ПДВП в соответствии с заданными стоимостными функциями.

Основываясь на выше сказанном, можно сделать вывод, что стратегия выбора структуры дерева ПДВП должна базироваться под заданный вычислительный ресурс (вычислительный ресурс – реконфигурируемая вычислительная среда: поточные, параллельно-поточные структуры процессоров); декомпозиция дерева преобразования осуществляется в процессе работы ПДВП-кодера, т.е. «на ходу». Преимущество данного подхода в том, что кодер учитывает локальные изменения статистики звукового сигнала. Любая структура дерева E может быть выражена окончательными его узлами (l, n) :

$$E_j = \{(l, n): l = 1, \dots, L, n = 0, \dots, 2^l - 1\}, \quad (4.16)$$

где j – номер структуры дерева из множества E ; L – максимальное значение масштабного уровня j -ой структуры дерева ПДВП.

Множество структур E_j образуют пространство структур деревьев ПДВП $E = \{E_1, E_2, E_3, E_4, \dots, E_N\}$, где E_N структура дерева, описывающая полный ПДВП. В силу того, что в структуре дерева ПДВП каждый узел может является предком двух потомков (двух новых узлов) при переходе с уровня l на новый уровень $l + 1$, то структуры ПДВП зависимы друг от друга:

$$E_j \in E_{j+1}. \quad (4.17)$$

Пусть задана некая структура дерева ПДВП (см. рис. 4.10) $E_3 = \{l_2(n_4, n_5, n_6, n_7), l_3(n_{10}, n_{11})\}$. Здесь пунктирными линиями показаны элементы, которыми можно изменить структуру дерева ПДВП. Изменение структуры дерева ПДВП E_j хотя бы одним пунктирным элементом, порождает появление новой структуры E_{j+1} , где структура дерева E_{j+1} отличается от структуры E_j только на добавленный элемент. Из этого следует, что структура дерева E_{j+1} содержит структуру дерева E_j , причём условие принадлежности (4.17) выполняется.

Следовательно, здесь возможна оптимизация декомпозиции дерева на основе методов сравнения [11]. Данный подход заключается в том, что лучшая структура дерева преобразования $(l, n) \in E_j$ с M частотными полосами может быть получена на основе добавления экстра ячеек (двух ветвей фильтров низких и высоких частот) к лучшей структуре дерева ПДВП $(l, n) \in E_{j-1}$ с $M - 1$ полосами. Таким образом, всё начинается с дерева на одной ячейке, далее могут делиться h_0 и h_1 полосы, т.е. ищется лучшее дерево на двух или трёх ячейках и выбор делается на основе сравнения двух результирующих банков фильтров.

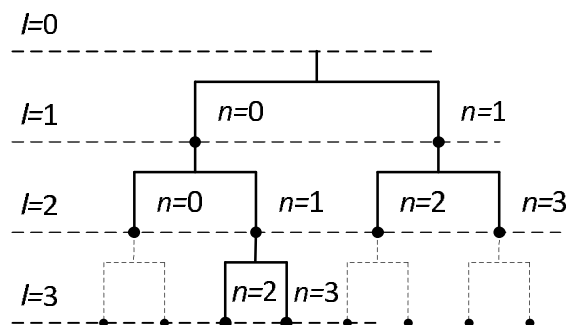


Рис. 4.10 – Пример структуры ПДВП, описанной через узлы (l, n) структуры дерева E_j ПДВП

Поиск лучшей структуры дерева преобразования не означает, что делается деление на полосы частотного диапазона («рост» дерева ПДВП) до последнего масштабного уровня. Это осуществляется на основе некоего критерия. При этом, декомпозиция дерева ПДВП выбирается как можно ближе к шкале барков [11,12].

Предположим, что необходимый вычислительный ресурс, под которым понимается производительность процессора и ёмкость его памяти, для одной ячейки преобразования (зависит от длины импульсной характеристики вейвлет фильтров) на фрейме входных данных есть c . Предположим также, что имеющийся в распоряжении вычислительный ресурс равен C . Таким образом, проблема заключается в адаптивном построении структуры дерева ПДВП $(l, n) \in E_j, j = 1, 2, 3, \dots$, при которой достигается минимум стоимостной функции J для заданного вычислительного ресурса C и временного сегмента звукового сигнала длиной, кратной степени двух, т.е. конструируется задача минимизации:

$$\min J(X_{l,n,k}(t), (l, n) \in E_j, k \in Z) \text{ для } c_j < C, \quad (4.18)$$

где c_j – требуемый вычислительный ресурс j -го дерева ПДВП.

Принимая во внимание тот факт, что вычислительная сложность полного дерева ПДВП равна

$$O(WPD) = 2P \cdot \sum_{l=0}^{L-1} 2^l \approx 4P, \quad (4.19)$$

или другими словами определяется вычислительной сложностью ячеек на первом масштабном уровне и на одном из l -х масштабных уровней (благодаря децимации 2:1 на каждом масштабном уровне), производительность каждой ступени поточного процессора выбирается одинаковой. В то время, как ресурс памяти на ступенях процессора и их число задаются предельной структурой преобразования: $CB - WPD: (l, n) \in E_{CB}$.

4.2.2 Выбор фильтров прототипов для ПДВП

Важным аспектом для анализа сигнала на основе ПДВП является выбор фильтров прототипов для банка фильтров [4]. На рис. 4.11 иллюстрирует пример амплитудно-частотных характеристик банков фильтров в зависимости от номера материнской вейвлет функции из семейства функций Даубеши (Daubechies).

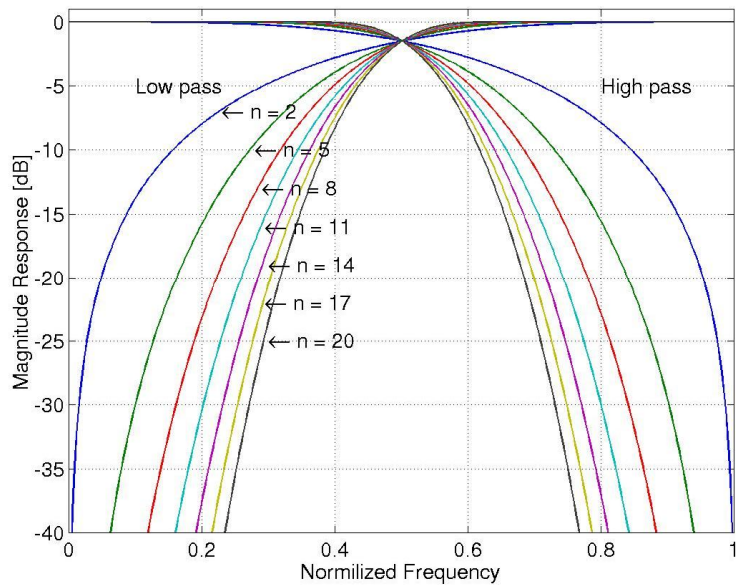


Рис. 4.11 – Амплитудно-частотная характеристика банка фильтров

На рис. 4.12 показана масштабная и вейвлет функция для фильтров прототипов из семейства материнских вейвлет функций Даубеши 20 порядка (40 коэффициентов).

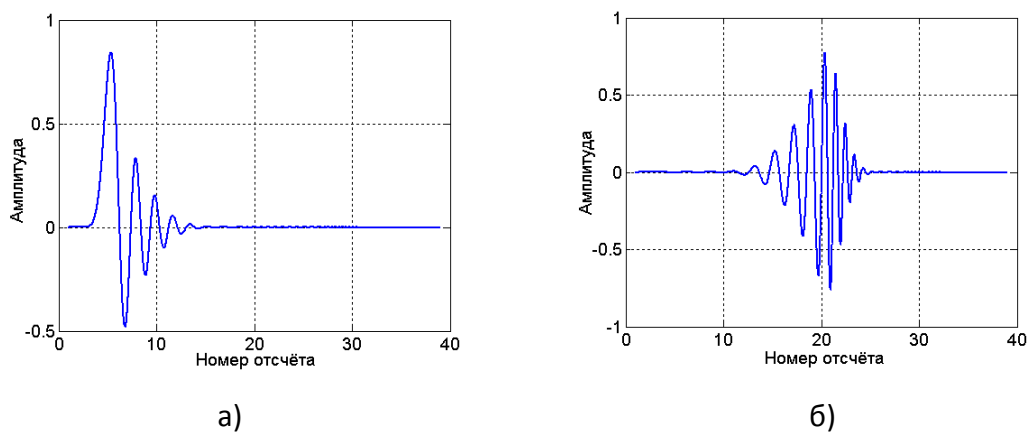


Рис. 4.12 – Функция Добеши 20-го порядка: а) масштабная; б) вейвлет функция

Значения параметра n изменяются в пределах от 2 до 20 и определяют длину выбранного фильтра прототипа, как $2n$. Выбор материнской функции диктуется конечным применением ПДВП, позволяет варьировать избирательность фильтров нижних и верхних частот, а так же количество коэффициентов, что определяет вычислительную сложность реализации.

Программа вычисления коэффициентов вейвлет и масштабной функций.

```
function [out1,out2,out3,out4] = wavefilters(wname,o)
%WAVEFILTERS Wavelet filters.
%
% [LO_D,HI_D,LO_R,HI_R] = WAVEFILTERS('wname') computes four
% filters associated with the orthogonal or biorthogonal
% wavelet named in the string 'wname'.
% The four output filters are:
% LO_D, the decomposition low-pass filter
% HI_D, the decomposition high-pass filter
```



```

%           LO_R, the reconstruction low-pass filter
%           HI_R, the reconstruction high-pass filter
%   Available wavelet names 'wname' are:
%   Daubechies   : 'db1' or 'haar', 'db2', ... , 'db50'
%
%   [F1,F2] = WAVEFILTERS('wname','type') returns the following
%   filters:
%   LO_D and HI_D if 'type' = 'd' (Decomposition filters)
%   LO_R and HI_R if 'type' = 'r' (Reconstruction filters)
%   LO_D and LO_R if 'type' = 'l' (Low-pass filters)
%   HI_D and HI_R if 'type' = 'h' (High-pass filters)

wname      = lower(wname);

% orth. wavelet
F = dbwavf(wname);
[LoF_D,HiF_D,LoF_R,HiF_R] = orthfilt(F);

if nargin==1
    out1 = LoF_D; out2 = HiF_D; out3 = LoF_R; out4 = HiF_R;
else
    o = lower(o(1));
    switch o
        case 'd' , out1 = LoF_D; out2 = HiF_D;
        case 'r' , out1 = LoF_R; out2 = HiF_R;
        case 'l' , out1 = LoF_D; out2 = LoF_R;
        case 'h' , out1 = HiF_D; out2 = HiF_R;
        otherwise
            error(['invalid argument value' msg]);
    end
end
end

```

Программа вычисления вейвлет коэффициентов функции Даубеши от 1 по 10 при выполнении операций декомпозиции и восстановления.

```

function F = dbwavf(wname)
%DBWAVEF Daubechies wavelet filters.
%   F = DBWAVEF(W) returns the scaling filter
%   associated with Daubechies wavelet specified
%   by the string W, where W = 'dbN'.
%   Possible values for N are:
%   N = 1, 2, 3, ... , 50.

if lower(wname(1:2))== 'ha' num = 1;
else
    lw = length(wname); ab = abs(wname);
    ii = lw+1;
    while (47<ab(ii-1)) & (ab(ii-1)<58) , ii = ii-1; end
    num = str2num(wname(ii:lw));
end
if num==1
    F = [ 0.500000000000000 0.500000000000000 ];

```

```

elseif num==2
    F = [ 0.34150635094622  0.59150635094587 ...
          0.15849364905378 -0.09150635094587 ];
elseif num==3
    F = [ 0.23523360389270  0.57055845791731  0.32518250026371 ...
          -0.09546720778426 -0.06041610415535  0.02490874986589 ];
elseif num==4
    F = [ 0.16290171402562  0.50547285754565  0.44610006912319 ...
          -0.01978751311791 -0.13225358368437  0.02180815023739 ...
          0.02325180053556 -0.00749349466513 ];
elseif num==5
    F = [ 0.11320949129173  0.42697177135271  0.51216347213016 ...
          0.09788348067375 -0.17132835769133 -0.02280056594205 ...
          0.05485132932108 -0.00441340005433 -0.00889593505093 ...
          0.00235871396920 ];
elseif num==6
    F = [ 0.07887121600143  0.34975190703757  0.53113187994121 ...
          0.22291566146505 -0.15999329944587 -0.09175903203003 ...
          0.06894404648720  0.01946160485396 -0.02233187416548 ...
          0.00039162557603  0.00337803118151 -0.00076176690258 ];
elseif num==7
    F = [ 0.05504971537285  0.28039564181304  0.51557424581833 ...
          0.33218624110566 -0.10175691123173 -0.15841750564054 ...
          0.05042323250485  0.05700172257986 -0.02689122629486 ...
          -0.01171997078235  0.00887489618962  0.00030375749776 ...
          -0.00127395235906  0.00025011342658 ];
elseif num==8
    F = [ 0.03847781105406  0.22123362357624  0.47774307521438 ...
          0.41390826621166 -0.01119286766665 -0.20082931639111 ...
          0.00033409704628  0.09103817842345 -0.01228195052300 ...
          -0.03117510332533  0.00988607964808  0.00618442240954 ...
          -0.00344385962813 -0.00027700227421  0.00047761485533 ...
          -0.00008306863060 ];
elseif num==9
    F = [ 0.02692517479416  0.17241715192471  0.42767453217028 ...
          0.46477285717278  0.09418477475112 -0.20737588089628 ...
          -0.06847677451090  0.10503417113714  0.02172633772990 ...
          -0.04782363205882  0.00017744640673  0.01581208292614 ...
          -0.00333981011324 -0.00302748028715  0.00130648364018 ...
          0.00016290733601 -0.00017816487955  0.00002782275679 ];
elseif num==10
    F = [ 0.01885857879640  0.13306109139687  0.37278753574266 ...
          0.48681405536610  0.19881887088440 -0.17666810089647 ...
          -0.13855493935993  0.09006372426666  0.06580149355070 ...
          -0.05048328559801 -0.02082962404385  0.02348490704841 ...
          0.00255021848393 -0.00758950116768  0.00098666268244 ...
          0.00140884329496 -0.00048497391996 -0.00008235450295 ...
          0.00006617718320 -0.00000937920789 ];
else
    error('Invalid Daubechies wavelet number !');
end

```

4.2.3 Декомпозиция ПДВП на основе алгоритма «простого дерева»

Задача оптимизации структуры ПДВП для систем кодирования аудио сигналов, заключается в поиске субоптимальной структуры дерева ПДВП. Оптимум может быть найден основываясь на исходных параметрах системы кодирования, таких как: количество вносимых искажений и скорость передачи.

Алгоритм поиска лучшей структуры дерева ПДВП, через параметры системы кодирования: скорость передачи R и количество вносимых искажений D , в литературе [13,14] называется алгоритмом «простого дерева». Зависимость параметров R и D в структуре дерева ПДВП для каждого уровня декомпозиции можно записать, как

$$R_{l,n} = R_{l+1,2n} + R_{l+1,2n+1} \quad (4.20)$$

и

$$D_{l,n} = D_{l+1,2n} + D_{l+1,2n+1}. \quad (4.21)$$

где l и n определяют уровень декомпозиции и номер узла соответственно в структуре дерева ПДВП E . Лучшая структура ПДВП E_{opt} в данном случае зависит от взаимной зависимости параметров D и R . Соответственно, логично будет определить выражение отражающее зависимость количества вносимых искажений D при фиксированной скорости передачи R . Составим зависимость между параметрами D и R через множитель Лагранжа λ . Таким образом, следующее соотношение определяет стоимостную функцию Лагранжа как

$$J(\lambda) = D + \lambda R. \quad (4.22)$$

Задача поиска оптимального множителя сводится к нахождению минимального значения стоимостной функции Лагранжа, которая определяет дальнейшую декомпозицию узлов дерева ПДВП.

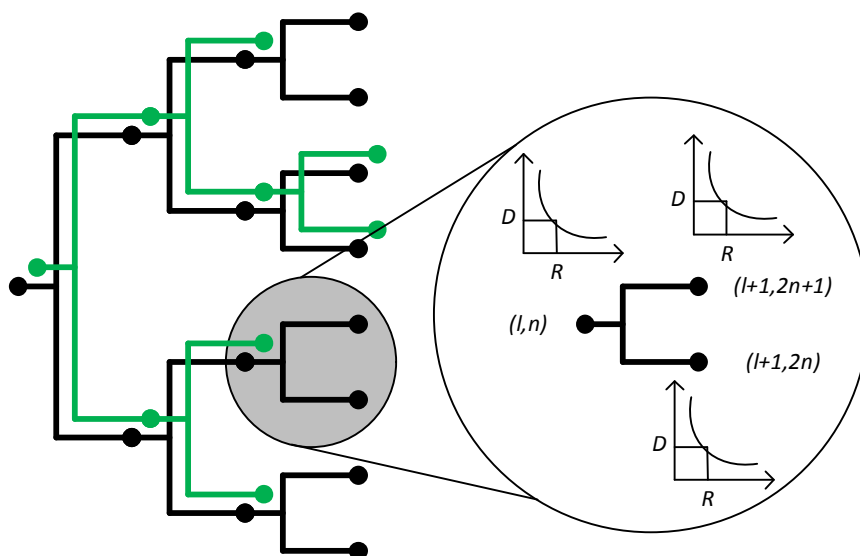


Рис. 4.13 – Алгоритм нахождения структуры адаптивного ПДВП на базе алгоритма «простого дерева»

Задачу нахождения стоимостной функции Лагранжа для полного дерева вейвлет преобразования можно представить в виде алгоритма:

Этап 1. Инициализация. Строится полное дерево пакета вейвлет преобразования (данное дерево показано на рис. 4.13 черным цветом) и определяются параметры D и R для каждой ветви дерева.

Этап 2. Выполняется оптимизация дерева для заданного множителя Лагранжа. Вычисляется стоимостная функция Лагранжа для каждой базовой декомпозиции (рис. 4.13 выделенное в круге) и минимизируется стоимостная функция $J(X_{l,n})$ на базе выбранного квантователя

$$J(X_{l,n}) = \min_{\text{квантователь}} [D(X_{l,n}) + \lambda R(X_{l,n})]. \quad (4.23)$$

Данные действия повторяются рекурсивно для каждой базовой декомпозиции и значения сравниваются в соответствии с процедурой роста дерева ПДВП. Процедура сравнения в алгоритме повторяется начиная с самого нижнего уровня дерева и до его вершины. Результатом процедур минимизации и упрощения становится оптимальное значение стоимостной функции Лагранжа, которое определяет значение множителя Лагранжа. Далее выполняется операция синтеза новой структуры дерева с учетом полученного значения множителя Лагранжа (рис. 4.13 дерево серого цвета), адаптированного под первоначально заданную скорость передачи R и минимальным количеством вносимых искажений D .

Первые два этапа нахождения лучшего дерева ПДВП основываются на знании первоначального значения множителя Лагранжа, на практике это известно только в априори. Значение множителя Лагранжа зависит от скорости передачи и качества кодирования.

Этап 3. Нахождение оптимального значения множителя Лагранжа для заданной скорости передачи. Оптимальное значение λ вычисляется через быстро восходящий градиентный метод.

Вычислительная сложность данного алгоритма поиска наилучшей структуры ПДВП, оценивается как $O(Nd)$, где N - длина входного сигнала, d - глубина дерева (базовая декомпозиция 2 (см. рис. 4.2)).

4.2.4 Декомпозиция ПДВП на основе алгоритма «двойного дерева»

Сигнал с изменяющимися временными характеристиками требует применения более гибкого преобразования для анализа. В алгоритме «двойного дерева» [14] выбор лучшей структуры дерева ПДВП производится из структур вычисленных для каждого сегмента входного сигнала. Следовательно, данный алгоритм представляет собой иерархическое расширение алгоритма «простого дерева».

Алгоритм наглядно проиллюстрирован на примере, смотри рис. 4.14, где сплошной линией показано полное двухуровневое дерево ПДВП, назовем его как частотная декомпозиция сигнала (frequency decomposition). Аналогично частотной декомпозиции сигнала строится пространственная декомпозиция сигнала (spetial segmentation), которая показана на рис. 4.14 пунктирной линией, а переменные $T(1)$, $T(2)$, $T(3)$, $T(4)$ определяют первую, вторую, третью и четвертую части сегментов входного сигнала соответственно. Далее, в соответствии с алгоритмом «двойного дерева» к частотной и пространственной декомпозиции сигнала применяется алгоритм «простого дерева» для полного T , половины $T(1,2)$ и $T(3,4)$, и четверти $T(1)$, $T(2)$, $T(3)$, $T(4)$ сигнала соответственно, для обеих частей бинарного дерева.

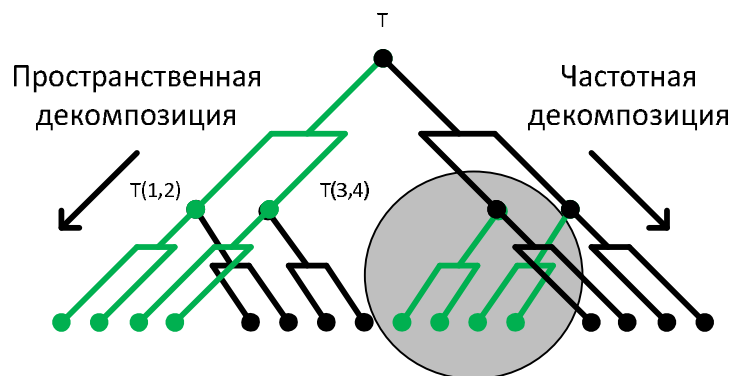


Рис. 4.14 – Алгоритм «двойного дерева» и алгоритм «частотно-временного дерева»

Результаты минимизации стоимостных функции Лагранжа каждой последовательности $J(T), J(T(1,2)), J(T(3,4)), J(T(1)), J(T(2)), J(T(3)), J(T(4))$ подставляются в пространственное дерево (на рис. 4.14 показаны две базовых декомпозиции черного цвета в пространственном дереве) и опять применяется алгоритм «простого дерева». Результат вычислений есть последовательность минимальных стоимостных функции для каждого сегмента (например для сегментов $T(1,2), T(3), T(4)$), определяемые значениями множителей Лагранжа, на основании которых синтезируются новые структуры деревьев ПДВП для каждого сегмента анализируемого сигнала.

Вычислительная сложность данного алгоритма значительно возрастает по сравнению с алгоритмом «простого дерева» и может быть оценена как $O(Nd^2)$, где N - длина входной последовательности, d – глубина дерева (базовая декомпозиция 2 (Рис. 4.2)).

4.2.5 Декомпозиция ПДВП на основе алгоритма «частотно-временного дерева»

Алгоритм «двойного дерева» представляет собой частный случая общего алгоритма «частотно-временного дерева»[14]. Суть алгоритма остается прежней, но структура генерируемого дерева анализа значительно расширяется за счет числа анализируемых ветвей. Оптимизация базовой декомпозиции выполняется на основе алгоритма «простого дерева». Алгоритм «частотно-временного дерева» осуществляет нахождение минимальных значений стоимостных функции Лагранжа для базовых декомпозиций дерева ПДВП и пространственного дерева (аналогично алгоритму «двойного дерева»). Минимальные значения стоимостных функций определяют ту часть структуру пространственного дерева, которая будет перенесена в частотное дерево и наоборот. Повторная оптимизация применяется для вновь сформированного дерева ПДВП. Минимизировав все стоимостные функции текущего уровня осуществляется переход на следующий уровень декомпозиции бинарного дерева (рис. 4.14 алгоритма «частотно-временного дерева»). Операции минимизации стоимостных функций для каждой части бинарного дерева повторяются до нахождения субоптимального минимума стоимостной функции. На основе полученного субоптимума, затем синтезируется новая структура дерева ПДВП для каждого временного сегмента входного сигнала. Сложность применения алгоритма «частотно-временного дерева» заключается в логарифмически возрастающей вычислительной сложности с увеличением глубины первоначального дерева. Однако, применение данного алгоритма по-

зволяют найти субоптимальное соотношение между выделенным сегментом сигнала и применяемой для его анализа структурой дерева ПДВП.

Оценить вычислительную сложность алгоритма «частотно-временного дерева» можно как $O(N2^d)$, где N - длина полного входного сигнала, d – глубина анализируемого дерева (базовая декомпозиция 2 (см. рис. 4.2)).

4.2.6 Декомпозиция ПДВП на основе перцептуальной энтропии

В перцептуальном кодировании звука, в первую очередь, интересна такая декомпозиция дерева ПДВП, при которой минимизируется скорость передачи с сохранением соответствующего высокого качества восприятия человеком декодированного сигнала. В данном случае, по мере обработки входного фрейма сигнала решение задачи предлагается разбить на два этапа:

- на основании стоимостных функций для каждого узла дерева преобразования $(l, n) \in E$, определяемых перцептуальной энтропией, осуществляется декомпозиция данных узлов при которой будет минимизировано требуемое количество бит кодирования вейвлет коэффициентов $X_{l,n}$, а шумы квантования не воспринимаются слухом человека. Таким образом, осуществляется рост дерева преобразования;
- оценивается информативность новой структуры дерева преобразования, т.е. как точно новое частотно-временное разрешение банка фильтров анализа локализовало нестационарности сигнала, а также требуемый вычислительный ресурс.

Стоимостная функция декомпозиции узлов $(l, n) \in E$ дерева ПДВП выбирается следующим образом:

$$J_{l,n} = \sum_{\forall X_{l,n,k}} \log_2(2[n \text{ int}(SMR_{l,n,k})] + 1), (l, n) \in E, k \in Z, \quad (4.24)$$

где $SMR_{l,n,k}$ – представляет собой отношение сигнала к порогу маскирования в полосе узла (l, n) дерева E . Другими словами, соотношение (4.24) вычисляется для каждой частотной полосы (узла дерева ПДВП) и представляет собой отношение среднеквадратического значения сигнала в узле (l, n) к среднеквадратическому значению шума квантования, который будет внесен в реконструированный сигнал. Максимально возможный уровень шума, не воспринимаемый на слух человеком, очевидно, является значением порога маскирования. Таким образом, весовая функция $J_{l,n}$ является индикатором необходимого числа битов для кодирования звукового сигнала. В качестве меры информативности ПДВП может быть выбрана энтропия [6]:

$$H(u) = \sum_k p(n) \log \frac{1}{p(n)}, \quad (4.25)$$

где $p(k) = |u(n)|^2 / \|u\|^2$ – нормализованная энергия k -го элемента вектора $u = \{u(n)\}, n = 1, 2, 3, \dots$ представленная функцией распределения вероятности p , причем $p \log(1/p) = 0$ для $p = 0$.

Величина энтропии показывает ту информативность, которая заключается в структуре дерева ПДВП. Другими словами, насколько сократилось разнообразие или неопределенность

описания входного сигнала деревом ПДВП. Применительно к кодированию звука на основе ПДВП-кодера можно судить, насколько точно удалось выбрать частотно-временное разрешение для локализации нестационарности в сигнале. Таким образом, мера качества информативности (мера неопределенности) $H(p_1, p_2, \dots, p_N)$, получается при полном устранении разнообразия и неопределенности в описании входного сигнала состояниями вейвлет коэффициентов дерева ПДВП, должна удовлетворять ряду условий: монотонное возрастание, причем недопустимые состояния не должны учитываться, так как они не меняют неопределенность; аддитивность (неопределенность объединенного источника должна равняться сумме неопределенностей исходных источников); непрерывность функции вероятности p_1, p_2, \dots, p_N , причём $\sum_i^N p_i = 1$; быть функцией от функции распределения и не зависеть от её конкретных значений, а также пути выбора вейвлет коэффициентов из множества структур ПДВП.

Неопределенность выбора вейвлет коэффициентов складывается из неопределенностей выбора вейвлет коэффициентов в каждой группе (дереве ПДВП), рассчитанных с учетом вероятностного выбора данной группы (данного дерева) (определяется алгоритмом роста дерева ПДВП):

$$\begin{aligned} H(p_1 p_2 \dots p_{N-2} q_1 q_2 q_3 q_4) \\ = H(p_1 p_2 \dots p_{N-2} p_{N-1} p_N) + p_{N-1} H(q_1/p_{N-1}, q_2/p_{N-1}) \\ + H(q_3/p_N, q_4/p_N), \end{aligned} \quad (4.26)$$

где q_1, q_2, q_3 и q_4 – вероятности состояния, образующих соответственно группы $N - 1$ и N , причем $p_{N-1} = q_1 + q_2$ и $p_N = q_3 + q_4$. Исходя из свойств энтропии [6], в частности, характеризующих среднюю неопределенность выбора, применительно в ПДВП-кодеру, предлагается конструировать меру количества информативности ПДВП (неопределенности) некой структуры дерева преобразования из множества структур в виде следующей стоимостной функции:

$$J_{l,n} = \sum_{\forall (l,n) \in E_j} |X_{l,n,k}|^2 \log \frac{1}{|X_{l,n,k}|^2}, (l,n) \in E, k \in Z. \quad (4.27)$$

С учетом определения энтропии (4.25) несложно показать, что

$$H(X_{l,n,k}) = \frac{J_{l,n}}{\|X_{l,n,k}\|^2} + \log \|X_{l,n,k}\|^2, (l,n) \in E, k \in Z, \quad (4.28)$$

т.е. минимизация стоимостной функции $J_{l,n}$ ведет к минимизации энтропии $H(X_{l,n,k})$, $(l,n) \in E_j$.

4.3 Согласованная подгонка на основе адаптивного ПДВП

Алгоритм «согласованной подгонки» (Matching Pursuit (MP)) на основе вейвлет анализа позволяет выполнить разложение входного сигнала по вейвлет функциям, частотные характеристики которых имитируют соответствующие характеристики сигнала. Данные функции формируют словарь атомов (ортонормальных векторов декомпозиции) на основе ПДВП, каждая из

которых хорошо локализована как по частоте так и масштабу. Полное дерево ПДВП позволяет сформировать избыточный словарь векторов декомпозиции.

В основе алгоритма МР лежит аппроксимация анализируемого сигнала $x(n)$ атомами g_γ выбираемыми из словаря D [15,16]. Для N отсчетов входного сигнала словарь D содержит N векторов $g_\gamma \in D$, где $\gamma = (l, n, k)$ индекс вектора для каждого уровня $0 < l < \log_2(N)$ узла $0 < n < 2^l$ и коэффициента $0 < k < 2^l N$ декомпозиции. Каждый вектор имеет время-частотную локализацию схожую с дискретной функцией окна, которая увеличиваясь на 2^l и имеет центр в $2^l(k + 1/2)$.

Реализация алгоритма МР заключается в итерационном повторении следующих шагов. На первой шаге, алгоритма МР выполняет декомпозицию сигнала $x(n)$ банком фильтров на основе полного дерева ПДВП E_{full} . На втором шаге, выбирается наиболее значимый вейвлет коэффициент X_γ , т.е. коэффициент с абсолютным максимальным значением весового коэффициента a_γ . Ему в соответствие ставится атом g_γ из словаря D . Формирование результирующего вектора декомпозиции выполняется при помощи обратного ПДВП преобразования на основании структуры дерева ПДВП E_{full} . Результирующий вектор, вычитается из исходного сигнала $x(n)$ для получения остаточного сигнала $r(n)$ (см. рис. 4.15). Процесс повторяется, где для последующей итерации алгоритма входным сигналом будет остаточный сигнал с предыдущей итерации. На m -той итерации остаточный сигнал $r^m(n)$ вычисляется как:

$$r^m(n) = \begin{cases} x(n) & m = 0 \\ r^{m+1}(n) + a_\gamma^m \cdot g_\gamma^m(n) & m \neq 0' \end{cases} \quad (4.29)$$

где a_γ^m весовой коэффициент оптимального вектора $g_\gamma^m(n)$ на m -той итерации, а γ индекс словаря D на m -той итерации. Оптимальным вектором считается тот вектор, у которого получается наибольшее значение произведения с остаточным сигналом $\langle r^m, g_\gamma^m \rangle$ [17].

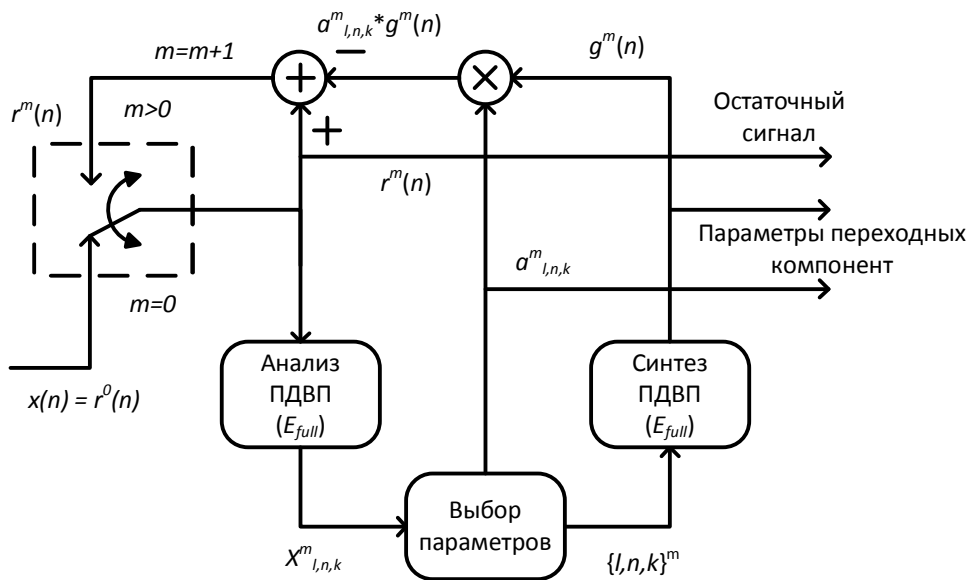


Рис. 4.15 – Структура алгоритма МР со словарем, построенным на основе полного дерева ПДВП

Алгоритм МР требует значительных вычислительных затрат от итерации к итерации при формировании субоптимального решения, которое в конце концов может быть не приме-

нимо в некоторых приложениях обработки звука и речи. Задача поиска оптимального словаря D построенного на основе ПДВП сводится к поиску лучшей структуры дерева декомпозиции ПДВП, как это отмечено в [13].

Задача моделирования переходных компонент звука решается на основе применения алгоритма МР. Формирование оптимального набора атомов в словаре D к входному сигналу, будет выполняться на основе применения адаптированного к кадру входного сигнала ПДВП, что позволит уменьшить размер формируемого словаря и сделать его динамическим, т.е. зависимым от структуры переходной компоненты сигнала. Таким образом структура системы моделирования переходных компонент сигнала будет состоять из двух частей. В первой части – оптимизация структуры дерева ПДВП под кадр входного сигнала на основе стоимостных функций. Во второй части – алгоритм МР с уже сформированным словарем, построенным на основе текущего ПДВП. Данный подход позволит уменьшить число итераций алгоритма МР и производить выбор только значимых компонент сигнала, что будет соответствовать оптимальным векторам декомпозиции.

Оптимизация структуры дерева ПДВП производится согласно описанному алгоритму анализа входного сигнала на основе адаптивного ПДВП, как это было описано выше, а так же в [18]. Структурная схема реализации алгоритма нахождения оптимальной декомпозиции входного сигнала ПДВП показана на рис. 4.16.

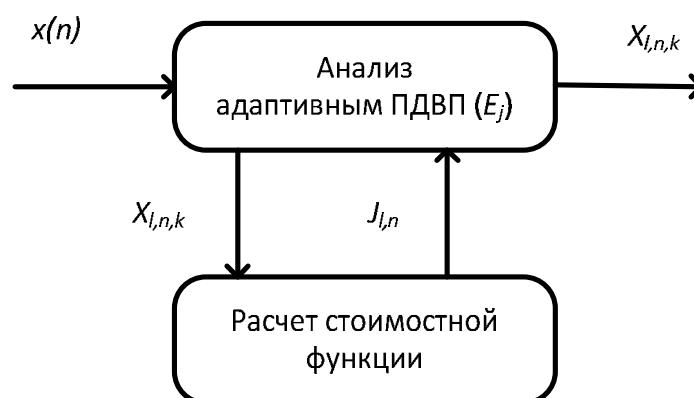


Рис. 4.16 – Структурна системы декомпозиции входного сигнала на основа адаптивного ПДВП

Поиск наилучшей аппроксимации анализируемого входного сигнала $x(n)$ векторами $g_{l,n,k}$ из словаря D , построенного на основе адаптированного ПДВП, сложная в вычислительном плане задача. Число атомов словаря D находится в прямой зависимости от число уровней ПДВП и длинны кадра сигнала. С точки зрения, например, аудиторной системы слуха человека, наиболее критичные компоненты сигнала расположены в области нижних, нежели в области верхних частот, что сказывается на выбор перцептуально значимых полос для анализа.

4.4 Список использованных источников

1. Vetterli M. and Kovacevic J., Wavelets and Subband coding, Englewood Cliffs. NJ : Prentice-Hall, 1995. p. 488.

-
2. Burrus C.S., Gopinath R.A. and Guo H., Introduction to wavelets and wavelet transforms, N.J. : Prentice Hall, 1998. p. 298.
 3. Воробьёв В.И. и Грибунин В.Г., Теория и практика вейвлет преобразования, С.-Петербург : Военный университет связи, 1999. стр. 203.
 4. Daubechies I., Ten lectures on Wavelets, Philadelphia, Pennsylvania : Society for industrial and applied mathematics, 1992. p. 357.
 5. Чуи К., Введение в вейвлеты, Пер с англ. Жилейкина Я.М., Москва : Мир, 2001, стр. 412.
 6. Wickerhauser M.V., Adaptive Wavelet Analysis from Theory to Software. Massachusetts : A.K. Peters Ltd., 1994. p. 486.
 7. Cohen I., Raz S. and Malah D., Orthonormal shift-invariant adaptive packet decomposition and representation, Signal Processing, Vol. 57, 3, March 1997, pp. 251-270.
 8. Петровский Ал., Анализ сигналов на основе пакета дискретного вейвлет преобразования с переменным временным шагом, Сборник трудов 6-й Международной научной конференции “Цифровая обработка сигналов и её применение” (DSPA’2004), Москва, Россия, 2004. Т. 1, стр. 186-189.
 9. Petrovsky Al. and al., et., Real-Time Wavelet Packet-based Low Bit Rate Audio Coding on a Dynamic Reconfigurable System, Amsterdam, Hetherlands : 114th AES Convention, 2003. Preprint 5778., 22 pages.
 10. Petrovsky Al., Masking threshold adapted wavelet packet (WP) tree with a time-invariant analysis wavelet, The Proc. of 6th International Conference “Pattern Recognition and Information Processing” (PRIP’2001). Minsk, Belarus, 2001. pp. 10-14.
 11. Philippe P., Saint-Martin F.M. and Lever M. Wavelet packet filterbanks for low time delay audio coding, IEEE Trans. Speech Audio Processing, 1999, Vol. 7, 3, pp. 310-322.
 12. Srinivasan P. and Jamieson L., High quality audio compression using an adaptive wavelet packet decomposition and psychoacoustic modeling, IEEE Trans. Signal Processing, Apr. 1998. vol. 46, 1998. pp. 1085-1093.
 13. Coifman R., Wickerhauser M.V., Entropy-Based Algorithms for Best Basis Selection, IEEE Trans. on Information Theory, Vol. 38, No. 2, March 1992 pp. 713-718
 14. Ramchandran K., Vetterli M., Best Wavelet Packet Bases in a Rate-Distortion Sense, IEEE Trans. on Image Processing, Vol. 2, No. 2, April 1993, pp. 160-175
 15. Mallat S. and Zhang Z., Matching pursuits with time-frequency dictionaries. IEEE Trans. on Signal Processing. December 1993, Vol. 41, 12, pp. 3397–3415.
 16. Vera-Candeas P., Ruiz-Reyes N. and Roza-Zurera M., Transient modelling by Matching-Pursuits with a wavelet dictionary for parametric audio coding, IEEE Signal Processing Letters, March 2004, Vol. 11, 3, pp. 349-352.

-
17. Vera-Candeas P., et al., New matching pursuit based sinusoidal modelling method for audio coding, IEE Proc. Vis. Image Signal Process, February 2004, Vol. 151, 1, pp. 21-28.
 18. Petrovsky Al., et al., Auditory modeling via frequency warped transforms. Amsterdam, Netherlands : Proc. of the 124th AES Convention, 2008. Preprint 8976, 15 pages.

5 ДИСКРЕТНОЕ ПРЕОБРАЗОВАНИЕ ФУРЬЕ С НЕРАВНОМЕРНЫМ ЧАСТОТНЫМ РАЗРЕШЕНИЕМ: ОПРЕДЕЛЕНИЕ, СВОЙСТВА, ВЫЧИСЛИТЕЛЬНЫЕ АСПЕКТЫ

Борович А., Парфенюк М., Петровский А.А.

АННОТАЦИЯ

В контексте перцептуальной обработки речевых сигналов в главе рассматривается ДПФ с неравномерным частотным разрешением (от англ. Warped discrete Fourier transform (WDFT)). Коэффициенты WDFT преобразования равномерно распределены в критических частотных полосах, поэтому в психоакустической модели на базе WDFT могут быть уравновешены как хорошее частотное, так и временное разрешение. Разработан метод минимизации ошибки реконструкции сигнала (вычисления обратного WDFT). Предложен алгоритм построения данного преобразования с избыточным базисом, обеспечивающим требуемую точность решения задачи синтеза.

Работа частично финансировалась по гранту W/WI/06/09 Белостотским политехническим институтом.

Ключевые слова - ДПФ, WDFT, фазовое звено, матрица Вандермонда, разложение по сингулярным числам матрицы.

СОДЕРЖАНИЕ

5.1	Введение.....	99
5.2	Определение WDFT.....	100
5.3	Обратное преобразование WDFT.....	101
5.3.1	Зависимость числа обусловленности матрицы преобразования от коэффициента деформации.....	101
5.3.2	Вычисление обратной матрицы.....	102
5.4	WDFT с избыточным синусоидальным базисом.....	104
5.4.1	Оценка ошибки реконструкции сигнала.....	104
5.4.2	Формирование избыточного синусоидального базиса.....	106
5.5	Вычислительные аспекты WDFT.....	110
5.6	Аппроксимация шкалы барков.....	111
5.7	Комментарии о частотном разрешении WDFT.....	112
5.8	Список использованных источников.....	114

5.1 Введение

ДПФ является мощным средством для частотного анализа с постоянной разрешающей способностью. Однако, в контексте перцептуальной обработки сигналов, основанной на психоакустической модели восприятия акустической информации человеком, возникает необхо-

димось в неравнополосной частотной декомпозиции сигнала в соответствии со шкалой критических частотных полос (Барков, ERB или MEL) [1,2]. Примером могут служить системы перцептуального кодирования звука и речи, а также подавления шума окружающей среды в речевом сигнале [3,4].

Идея построения преобразования с переменным частотным разрешением не нова. Впервые она была предложена почти 40 лет назад [5]. В настоящее время развиваются ряд новых нестандартных подходов [6-8]. Одним из них является дискретное преобразование Фурье с неравномерным частотным разрешением (от англ. Warped discrete Fourier transform (WDFT)) [8], позволяющее получить z -преобразование конечной последовательности входных отсчетов сигнала с неравномерным разложением коэффициентов преобразования на единичной окружности z -плоскости строго регулярно посредством использования фазового звена.

5.2 Определение WDFT

ДПФ с переменным частотным разрешением (NDFT) является наиболее обобщающим вариантом ДПФ [6]. Кроме единственности решения, NDFT не ограничивает расположение коэффициентов преобразования в z -плоскости. Дискретное преобразование Фурье с неравномерным разложением коэффициентов по частотной оси (WDFT) есть специальный случай NDFT, коэффициенты преобразования которого располагаются неравномерно, но строго регулярно на единичной окружности z -плоскости.

WDFT последовательности $x(n)$ из N точек определяется по следующей формуле:

$$\hat{X}(z_k) = X(\hat{z}_k) = \sum_{n=0}^{N-1} x(n) \hat{z}_k^{-n}, k = 0, \dots, N-1, \quad (5.1)$$

где \hat{z}_k -изображения, преобразованных фазовым звеном $A(z)$, равноотстоящих точек на единичной окружности в z -плоскости, получаемые из преобразования

$$\hat{z}_k^{-1} = e^{-j\frac{2\pi k}{N}} \rightarrow \hat{z}_k^{-1} = A(z_k), k = 0, \dots, N-1, \quad (5.2)$$

$A(z)$ устойчивое фазовое звено произвольного порядка.

Простейший вариант WDFT основан на фазовом звене первого порядка с действительным коэффициентом a

$$A(z) = \frac{z^{-1} - a}{1 - az^{-1}}. \quad (5.3)$$

Условием стабильности является $|a| < 1$. В зависимости от знака a , растягивается низкочастотный ($a > 0$) или высокочастотный ($a < 0$) диапазон, таким образом, оставшаяся часть единичной окружности в z -плоскости становится сжатой. Формально это может быть выражено следующим образом:

$$\hat{\omega} = \omega + 2 \arctan \left(\frac{a \cdot \sin(\omega)}{1 - a \cdot \cos(\omega)} \right) \text{ для } \begin{cases} z = e^{j\omega} \\ \hat{z} = e^{j\hat{\omega}} \end{cases} \quad (5.4)$$

то есть осуществляется билинейное преобразование z -плоскости в новую искривленную \hat{z} -плоскость (см. рис. 5.1).

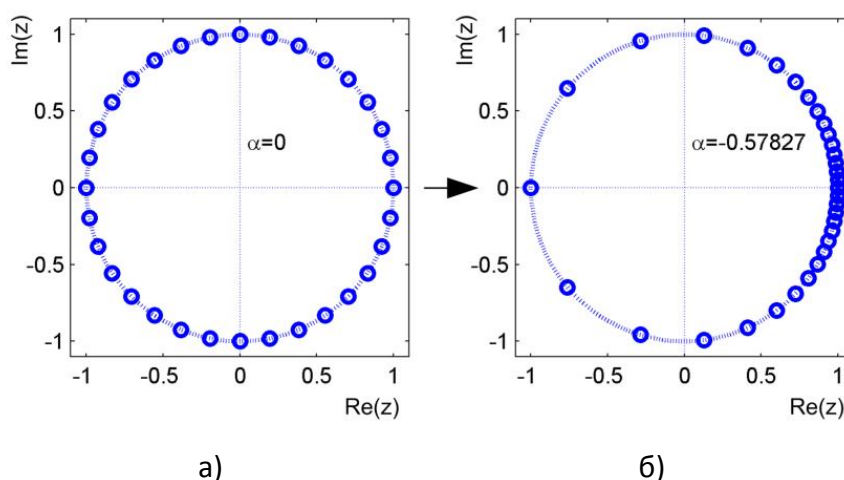


Рис. 5.1 – Расположение коэффициентов преобразования на единичной окружности в z -плоскости: ДПФ (а) и WDFT (б)

Как обобщение ДПФ, WDFT также имеет свойства линейности, симметрии и сдвига [7,8]. Сопряженная симметрия для действительных данных имеет силу и для WDFT:

$$\hat{X}(z_{N-1-k}) = \hat{X}^*(z_k), \quad (5.5)$$

однако ряд важных свойств ДПФ потеряно.

В матричной записи (с $\hat{X}(k)$, обозначая $\hat{X}(z_k)$) WDFT может быть представлено выражением:

$$\begin{bmatrix} \hat{X}(0) \\ \hat{X}(1) \\ \vdots \\ \hat{X}(N-1) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & A(z_0) & \dots & A(z_0)^{N-1} \\ 1 & A(z_1) & \dots & A(z_1)^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & A(z_{N-1}) & \dots & A(z_{N-1})^{N-1} \end{bmatrix}}_{\mathbf{D}} \cdot \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix}. \quad (5.6)$$

5.3 Обратное преобразование WDFT

5.3.1 Зависимость числа обусловленности матрицы преобразования от коэффициента деформации

В случае перцептуального деформирования оси частот матрицы Вандермонда, к классу которых относится и матрица \mathbf{D} WDFT, обладают свойствами, делающими невозможным получение точной обратной матрицы. А именно, матрица WDFT может рассматриваться как сингулярная так, как между некоторыми ее строками существуют почти линейные зависимости. Этот факт представляется как очень малое значение ее детерминанта $\det \mathbf{D} = \prod_{i < j} (\hat{z}_i^{-1} - \hat{z}_j^{-1})$ для определенного \hat{z}_k . Другими словами, WDFT-матрица является плохо обусловленной. Это означает, что любой численный алгоритм обращения матрицы, применяемый к подобной матрице, является очень чувствительным к малым изменениям данных.

Усиление ошибки может быть оценено, используя число обусловленности матрицы полного ранга (собственное значение матрицы)

$$\text{cond}(\mathbf{D}) = \|\mathbf{D}\| \|\mathbf{D}^{-1}\| = \sigma_{\max} / \sigma_{\min}, \quad (5.7)$$

где $\|\cdot\|$ означает произвольную норму матрицы (в общем случае Евклидову); σ_{\max} и σ_{\min} – наибольшее и наименьшее сингулярные числа.

Как показано на рис. 5.2, число обусловленности преобразуемой матрицы зависит от ее размерности и величины коэффициента деформации α и имеет очень большие значения даже при малоразмерных преобразованиях, слегка отличающихся от ДПФ. Плохая обусловленность является неотъемлемым свойством матриц Вандермонда, связанных с реальными проблемами. Единственным исключением является случай отсутствия деформирования частотной оси, когда WDFT становится обычным ДПФ.

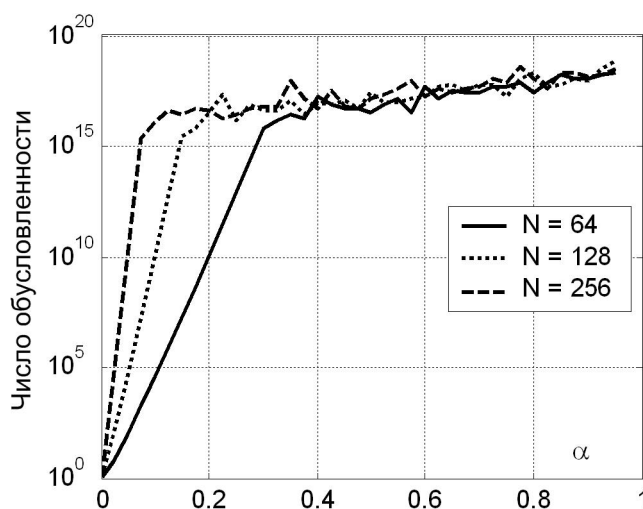


Рис. 5.2 – Зависимость числа обусловленности от коэффициента деформации α для различных размеров WDFT

5.3.2 Вычисление обратной матрицы

Существует ряд методов для аппроксимации обращений плохообусловленных матриц. Данные методы используют разложение по сингулярным числам матрицы (SVD) [9], которое определяется как

$$\mathbf{D} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H = \sum_{i=1}^N \mathbf{u}_i \sigma_i \mathbf{v}_i^H, \quad (5.8)$$

где

$$\begin{aligned} \mathbf{U} &= [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N], \\ \mathbf{V} &= [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N], \end{aligned} \quad (5.9)$$

матрицы с ортогональными столбцами $\mathbf{U}^H \mathbf{U} = \mathbf{V} \mathbf{V}^H = \mathbf{I}_N$ и

$$\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_N) \quad (5.10)$$

является диагональной матрицей состоящей из сингулярных чисел, отсортированных по убыванию $\sigma_1 \geq \dots \geq \sigma_N \geq 0$. Столбцы матриц \mathbf{U} и \mathbf{V} , левый и правый, – сингулярные векторы, H обозначает эрмитово транспонирование матрицы. Степень неполноты матрицы \mathbf{D} (наличие некоторых линейных зависимостей между столбцами) проявляется в существовании почти нулевых сингулярных чисел σ_i . Распределения сингулярных значений для некоторых различных матриц WDFT проиллюстрировано на рис. 5.3.

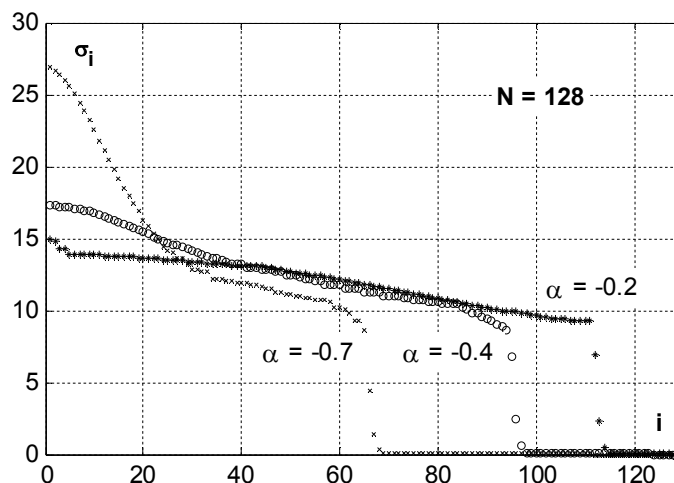


Рис. 5.3 – Распределение сингулярных чисел как функция коэффициента деформации α для фиксированного размера преобразования $N = 128$

Как показывает рис. 5.3, значения достаточно большого количества сингулярных чисел близки к нулю. Анализ распределения компонент SVD может дать много полезных пояснений плохой обусловленности матриц, а также, они могут быть использованы для формирования псевдообратной матрицы

$$\mathbf{D}' = \sum_{i=1}^N f_i \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^H. \quad (5.11)$$

Символы f_i обозначают так называемые коэффициенты ослабления фильтра и должны быть все равны единице для получения точного обращения матрицы. Теория регуляризации матриц [9] рекомендует исключать влияние малых сингулярных чисел, ослабляя их вклад в (5.11). Это реализуется путем установки соответствующих значений f_i . Так как сингулярные векторы с наибольшими индексами имеют ярко выраженный колебательный характер, то это действие изменяет спектральный состав данных.

Основное различие между известными методами регуляризации заключается в выборе коэффициентов ослабления фильтра. В простейшем подходе усеченного SVD (TSVD) [10] сумма (5.11) ограничивается обычным отбрасыванием термов имеющих сингулярные числа меньше определенного порога

$$f_i = \begin{cases} 1 & \sigma_i > \lambda, \\ 0 & \text{в противном случае.} \end{cases} \quad (5.12)$$

Менее радикальным является демпфированное SVD (DSVD), в котором коэффициенты ослабления фильтра постепенно изменяются в сторону нуля, обеспечивая более сглаженное отсечение

$$f_i = \frac{\sigma_i}{\sigma_i + \lambda}. \quad (5.13)$$

В обоих случаях, параметр регуляризации λ должен быть выбран с осторожностью. Решение должно быть стабильным при спектральных изменениях, ограниченных определенным минимумом.

5.4 WDFТ с избыточным синусоидальным базисом

5.4.1 Оценка ошибки реконструкции сигнала

В зависимости от приложения, где применяется WDFТ, ошибка реконструкции речевого сигнала (вычисление обратного WDFТ) может играть решающую роль на качество синтезированного сигнала (числе артефактов), например, в системах редактирования шумов окружающей среды в звуковом сигнале на основе психоакустически мотивированного правила взвешивания спектра зашумленного звука [11].

Используя матричную форму записи, вектор искажения сигнала \mathbf{d} можно определить как разность между оригинальным \mathbf{x} и реконструированным сигналами $\hat{\mathbf{x}}$:

$$\mathbf{d} = \mathbf{x} - \hat{\mathbf{x}} = (\mathbf{I} - \mathbf{D}'\mathbf{D})\mathbf{x}. \quad (5.14)$$

Соответствующая мера в частотной области может быть выражена как средняя спектральная плотность мощности (СПМ) искажения сигнала

$$S_{dd}(\omega) = \frac{1}{N} E\{|\mathbf{e}(\omega)^H \mathbf{d}|^2\} = \frac{1}{N} \mathbf{e}(\omega)^H \mathbf{R}_{dd} \mathbf{e}(\omega), \quad (5.15)$$

где

$$\mathbf{e}(\omega) = [1 \quad e^{-j\omega 1} \quad e^{-j\omega 2} \quad \dots \quad e^{-j\omega(N-1)}]^T, \quad (5.16)$$

является вектором столбцов синусоидального базиса ДПФ, а \mathbf{R}_{dd} обозначает ковариационную матрицу сигнала ошибки. Предположим, что входной сигнал \mathbf{x} является случайным вектором с нулевым математическим ожиданием и известной ковариационной матрицей $\mathbf{R}_{xx} = E\{\mathbf{x}\mathbf{x}^H\}$, тогда пусть $\mathbf{Q} = \mathbf{I} - \mathbf{D}'\mathbf{D}$, и матрица \mathbf{R}_{dd} может быть записана следующим образом:

$$\mathbf{R}_{dd} = \mathbf{Q}\mathbf{R}_{xx}\mathbf{Q}^H. \quad (5.17)$$

Очевидно, что спектральное искажение (5.15), зависит от характеристик входного сигнала и качества аппроксимации обратного WDFТ. Теоретически точное обращение матрицы \mathbf{D} возможно, но данное решение будет очень не стабильным вследствие большого числа обусловленности матрицы и не найдет практического применения в системах обработки звука. Однако если удастся получить стабильное решение обратной задачи, мера величины ошибки (5.15) может быть использована для регулирования величины артефактов в синтезированном речевом сигнале. Так, если псевдообратная матрица \mathbf{D}' определена с помощью техники SVD, представленной выше, то матрица \mathbf{Q} может быть вычислена напрямую из выражения

$$\mathbf{Q} = \mathbf{I} - \mathbf{D}'\mathbf{D} = \mathbf{I} - \sum_{i=1}^N f_i \mathbf{v}_i \mathbf{v}_i^H. \quad (5.18)$$

В случае перцептуальной деформации частотной оси, выбор параметров регуляризации ограничивает влияние на уровень спектральных искажений. Например, положим, что коэффициент ослабления фильтра f_i вычисляется по методу DSVD (5.13) с параметром регуляризации $\lambda = 0,001$. СПМ ошибки WDFT синтеза, вычисленная по формуле (5.15), показана на рис. 5.4. В качестве входного сигнала в эксперименте использовался стационарный окрашенный гауссовый шум, который моделировался с априори заданной теплицевой ковариационной матрицей $(\mathbf{R}_{xx})_{i,j} = p^{|i-j|}$ для $p = 0,9$. Можно заметить, что уровень искажений на заданной частоте зависит от расстояния между соседними WDFT коэффициентами. Сигнал полностью восстанавливается только в точках преобразования (частотах, определенных выбором коэффициента a фазового звена (5.3)) и спектральные искажения особенно заметны в растянутых частотных диапазонах, в то время как в сжатых частотных диапазонах ошибка синтеза имеет приемлемый уровень.

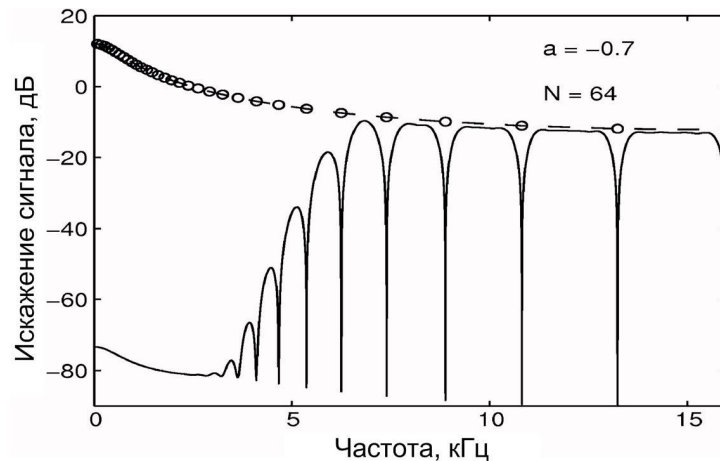


Рис. 5.4 – СПМ оригинального сигнала (кружочки) и соответствующая ошибка синтеза (сплошная линия)

Этот эффект наблюдается также и на спектрограмме реконструированного речевого сигнала в виде узких спектральных «дыр», локализованных между точками преобразования (см. рис. 5.5). Таким образом, частотная характеристика синтезированного речевого сигнала в некоторых деталях не восстанавливается. Заметим, что проявление данного эффекта в высокочастотной части частотного диапазона обусловлено перцептуальной деформацией частотной шкалы, т.е. коэффициент фазового звена ($a < 0$), но ситуация меняется и эффекты ошибки WDFT синтеза проявляются в низкочастотной части спектра, если ($a > 0$). Однако, данный случай не интересен для перцептуальных систем обработки речевых сигналов.

Известно, что высокочастотный диапазон спектра речевого сигнала влияет на качество речи, в то время, как низкочастотный – на разборчивость речи. Таким образом, при выборе параметров настройки WDFT необходимо искать некое компромиссное решение между требуемым частотным разрешением и ошибкой синтеза. Можно попытаться проектировать фазовое звено с коэффициентом a зависимым от времени, но данное решение имеет очень высокую

вычислительную сложность и более того, для широкополосного сигнала не возможно одновременно обеспечить его обработку с хорошей разрешающей способностью в низкочастотной части и перфективную реконструкцию его высокочастотных компонент, даже если удастся построить управляемое фазовое звено. Единственный путь минимизации спектрального искажения – это модификация матрицы преобразования \mathbf{D} таким образом, чтобы количество почти нулевых сингулярных чисел было уменьшено.

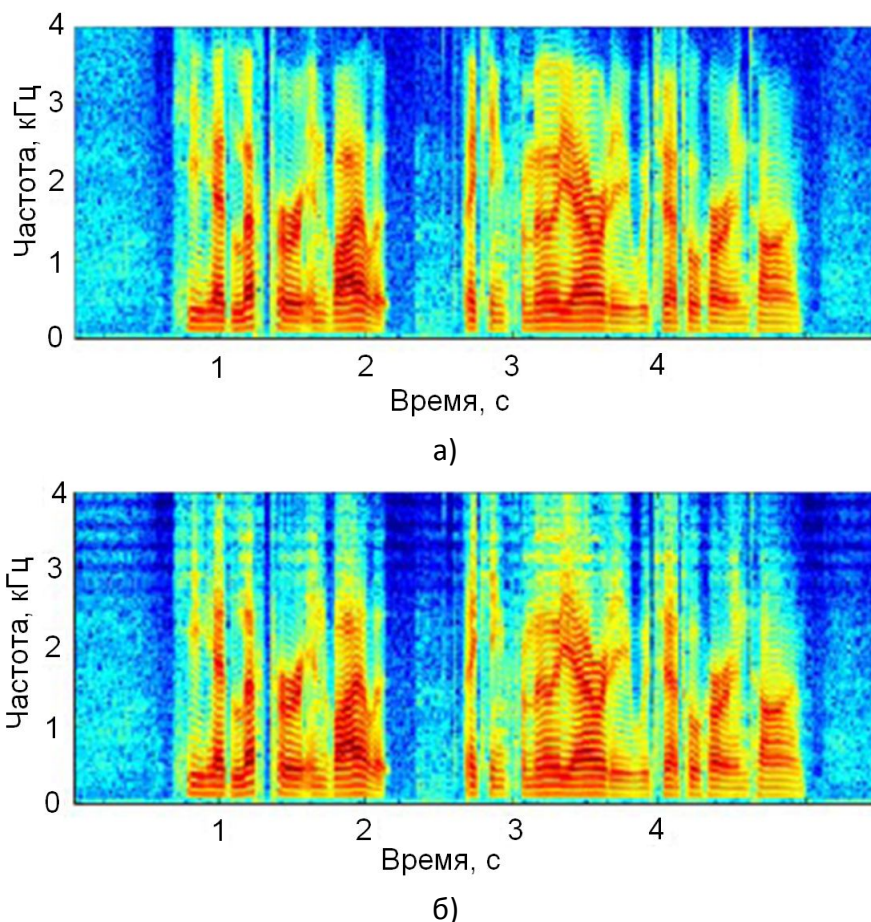


Рис. 5.5 – Спектрограммы оригинального речевого сигнала: (а) и реконструированного речевого сигнала (б)

5.4.2 Формирование избыточного синусоидального базиса

В работе [11] было показано, что формирование избыточного не ортогонального базиса, может внести значительный вклад в коррекцию ошибки реконструкции, а именно, при подходящем выборе векторов синусоидального базиса, можно модифицировать распределение сингулярных чисел для соответствующей матрицы преобразования. Новая WDFT матрица \mathbf{D} не является квадратной и количество строк $M > N$ увеличивается. Матричное представление WDFT с избыточным базисом может быть записано следующим образом:

$$\begin{bmatrix} \hat{X}(0) \\ \hat{X}(1) \\ \vdots \\ \hat{X}(N-1) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & \hat{z}_0^{-1} & \dots & \hat{z}_0^{-N+1} \\ 1 & \hat{z}_1^{-1} & \dots & \hat{z}_1^{-N+1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \hat{z}_{M-1}^{-1} & \dots & \hat{z}_{M-1}^{-N+1} \end{bmatrix}}_{\mathbf{D}_{M \times N}} \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix}, \quad (5.19)$$

Задача поиска обратной матрицы для данной прямоугольной матрицы $\mathbf{D}_{M \times N}$ [12,13] может быть также решена, используя SVD-процедуру, аналогичную как и для квадратной матрицы \mathbf{D} . Следует отметить, что стабильность новой матрицы выше, чем у матрицы стандартного WDFT. Избыточный базис с добавленными новыми векторами уменьшает эксцентricность SVD-эллипсоида, который является отображением единичной сферы в N -мерном пространстве. Соотношение между форматом WDFT с избыточным базисом и собственными числами для различных значений коэффициента деформации частотной шкалы a показаны на рис. 5.6. Видно, что если $M \rightarrow \infty$, то число обусловленности матрицы $\mathbf{D}_{M \times N}$ близко к единице.

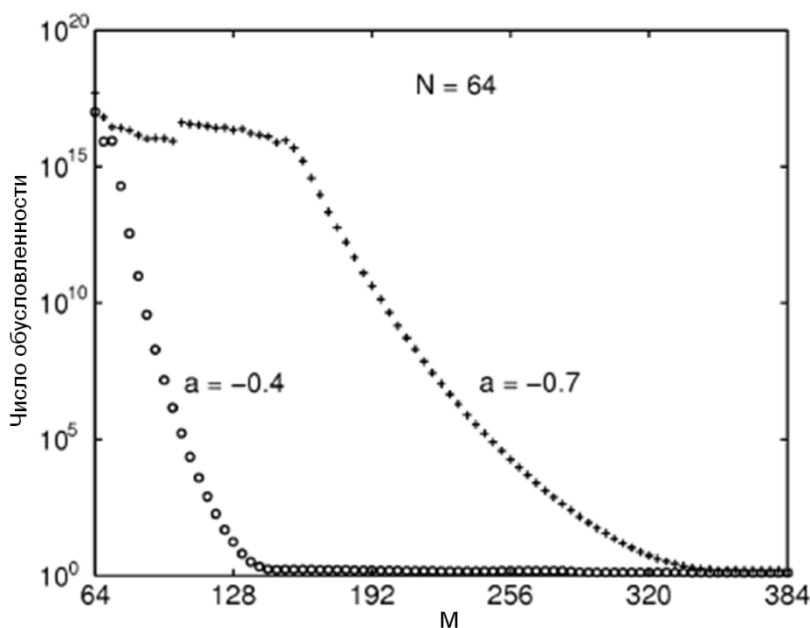


Рис. 5.6 – Соотношение между форматом WDFT с избыточным базисом и собственными числами матриц $\mathbf{D}_{M \times N}$

С практической точки зрения необходимо выполнить два условия при формировании избыточного базиса WDFT. Во-первых, сохранить регулярность расположения коэффициентов преобразования в соответствии с деформацией частотной шкалы (например, шкалой барков). Во-вторых, размер нового синусоидального базиса должен быть как можно меньшим, чтобы вычислительная сложность WDFT оставалась приемлемой.

Так как любое направление в пространстве комплексных векторов, определяемое базисным вектором, соответствует конкретному частотному диапазону, то WDFT можно представить как максимально децимированный банк фильтров. При этом, k -я строка WDFT матрицы может рассматриваться как КИХ-фильтр с передаточной функцией, определяемой по следующему выражению

$$H_k(z) = \sum_{n=-\infty}^{\infty} A(z_k)z^{-n}, k = 0, \dots, N - 1. \quad (5.20)$$

где $A(z)$ - фазовое звено первого порядка, $H_k(z)$ - полосовой фильтр с центральной частотой

$$\hat{\omega}_k = 2 \cdot \tan^{-1} \left(\frac{1+a}{1-a} \tan \left(\frac{\omega_k}{2} \right) \right), \omega_k = \text{angle}(z_k), \quad (5.21)$$

и полосой $2\pi/N$.

АЧХ фильтров $H_k(z)$ для $k = 0, \dots, N - 1$ представлены на рис. 5.7.

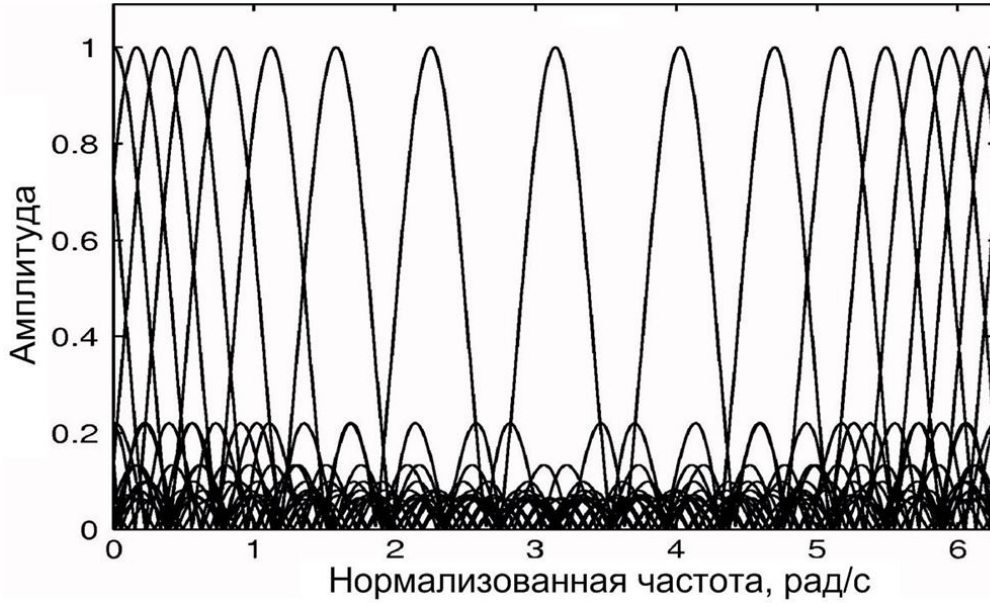


Рис. 5.7 – АЧХ критически децимированного банка фильтров WDFT

Новый избыточный базис WDFT должен быть составлен из M векторов, соответствующих импульсных характеристик КИХ-фильтров с центральными частотами

$$z_k = e^{j\pi k/M}, k = 0, \dots, M - 1, M > N, \quad (5.22)$$

регулярность которых на единичной окружности гарантирована, потому что вектор новых коэффициентов преобразования в точности совпадает со стандартным WDFT, но для формата преобразования M . Другими словами, размер нового избыточного базиса должен быть равен числу перекрывающихся полосовых фильтров (5.20) переопределенных для $k = 0, \dots, M - 1$. Если коэффициент фазового звена в (5.3) ($a < 0$), то максимальное расстояние между центральными частотами КИХ-фильтров составит:

$$\Delta\omega_{max} = \pi - 2 \tan^{-1} \left(\frac{1+a}{1-a} \tan \left(\frac{\pi - 2\pi/M}{2} \right) \right). \quad (5.23)$$

Для получения того же частотного разрешения в диапазоне верхних частот, что и у обычного ДПФ, расстояние между точками преобразования не должно быть больше чем $2\pi/N$. Подставляя $\Delta\omega_{max} = 2\pi/N$ в (5.21) и решая относительно M , получается, что

$$M = M_{opt} = 2\pi \left[\pi - 2 \tan^{-1} \left(\frac{1-a}{1+a} \tan \left(\frac{\pi - 2\pi/N}{2} \right) \right) \right]^{-1}. \quad (5.24)$$

Отметим, что для $a = 0$ (нет деформации частотной шкалы) $M = N$, следовательно, $M \geq N$. Соотношение между размером избыточного базиса M и коэффициентом деформации a для заданного числа столбцов матрицы $\mathbf{D}_{M \times N}$ WDFT с избыточным базисом иллюстрируется на рис. 5.8.

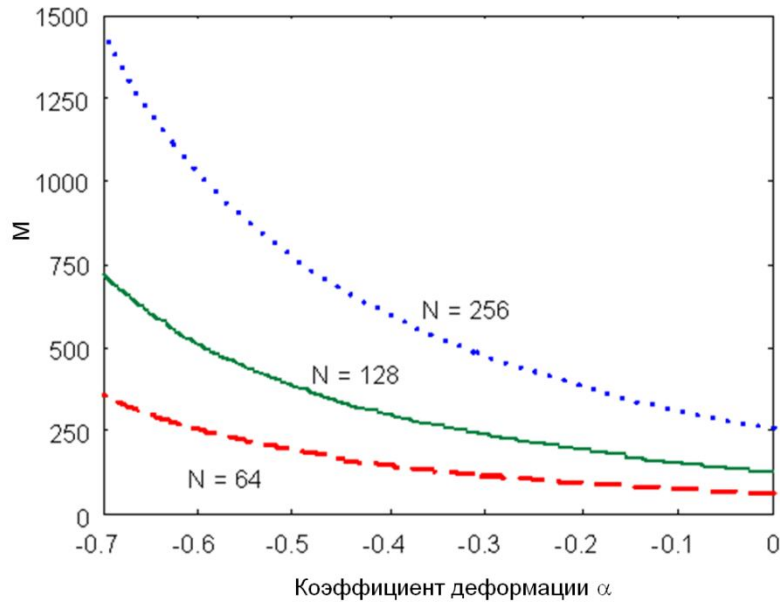


Рис. 5.8 – Соотношение между размером избыточного базиса и коэффициентом деформации α для заданного числа столбцов матрицы $\mathbf{D}_{M \times N}$

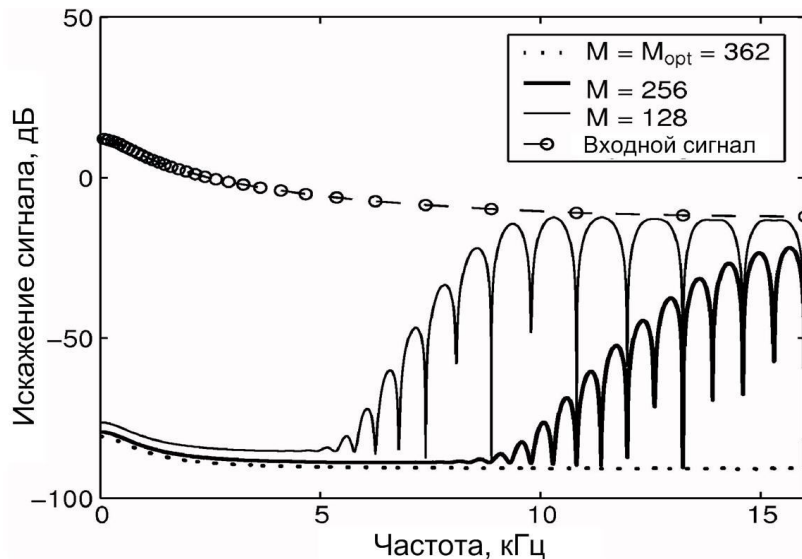


Рис. 5.9 – Спектральная плотность мощности входного сигнала и соответствующие спектральные искажения для разных размеров WDFT преобразования ($N = 64$)

Спектральные искажения (5.15) могут быть вычислены для прямоугольной матрицы также как и для квадратной, потому что продукт $\mathbf{D}'\mathbf{D}$ всегда $N \times N$ матрица. На рис. 5.9 показаны СПМ входного сигнала и соответствующей ошибки синтеза, вычисленные для прямоугольной матрицы $\mathbf{D}_{M \times N}$ WDFT. Как видно, ошибка синтеза уменьшается при увеличении M и может не учитываться при $M \geq M_{opt}$ (меньше -75 дБ). На практике число строк матрицы $\mathbf{D}_{M \times N}$ можно слегка уменьшить, так как перекрытие частотных характеристик полосовых фильтров (5.20) выбиралось случайно и без всякого психоакустического критерия. Однако прослушивание речевых тестов показало, что спектральные искажения не слышны в реконструированном сигнале речи для $M \approx M_{opt}$.

Ниже приведен листинг -программы формирования WDFТ-матрицы и вычисления псевдообратной WDFТ-матрицы на основе разложения по сингулярным числам матрицы.

```

function [W, W_1] = wdfmtx2(N,M,a,cond)
% generation of WDFТ2 matrix of size NxM with warping coefficient a

% [ e^(-jw[1]) e^(-jw[1]) ... e^(-jw[1]) ] .^ [ 0 1 ... M-1 ]
% [ e^(-jw[2]) e^(-jw[2]) ... e^(-jw[2]) ] [ 0 1 ... M-1 ]
% [ ... ... ... ] [ ... ... ... ]
% [ e^(-jw[N]) e^(-jw[N]) ... e^(-jw[N]) ] [ 0 1 ... M-1 ]

% z_1 = exp(-j * 2 * pi * (0 : (N - 1)) ./ N).';
% z_1 = (z_1 - a) ./ (1 - a .* z_1);
%
% W = (z_1*ones(1,M)).^(ones(N,1)*[0:M-1]);
% return;

N2 = ceil(N/2);
z_1 = exp(-j * 2 * pi * (0 : (N2 - 1)) ./ N).';
z_1 = (z_1 - a) ./ (1 - a .* z_1);

W = (z_1*ones(1,M)).^(ones(N2,1)*[0:M-1]);
if N/2 == N2,
    W = [ W; (-ones(1,M)).^[0:M-1]; conj(flipud(W(2:N2,:))) ];
else
    W = [ W; conj(flipud(W(2:N2,:))) ];
end;

if nargin < 2, return; end;

% generation of pseudoinverse matrix via damped SVD
% W = U*S*V', where U*U' = I and V*V' = I
%
% W_1 = V*S*U'
%
% S = diag( [ s[1]^(-1) s[2]^(-1) ... s[K]^(-1) 0 ... 0 ] ),
% where s[K] > lambda

[U, S, V] = svd(W, 'econ');
% perform reguarization
s = diag(S);
if nargin < 4, cond = 10; end;
lambda = max(s)/cond;
r = sum(s > lambda);
S_1 = diag([ones(r,1)./s(1:r); 0.0*ones(length(s)-r,1)]);
W_1 = V*S_1*U';

```

5.5 Вычислительные аспекты WDFТ

Алгоритм, по эффективности сравнимый с БПФ, не может быть построен для WDFТ из-за асимметричности WDFТ-матрицы. Тем не менее, прямой алгоритм работы с комплексной

матрицей может быть значительно оптимизирован. Наиболее усовершенствованный алгоритм, несмотря на сложность $O(N^2)$, использует факторизацию WDFT-матрицы в произведение действительной, ДПФ (осуществляемой через БПФ) и комплексной диагональной матрицы [8]. Данный метод хорошо подходит для обработки изображений, где все данные поступают одновременно. В обработке звука, где отсчеты поступают последовательно один за другим, может быть использована даже прямая реализация преобразования WDFT. Учитывая выражение (5.5), формула (5.6) может быть записана в следующем виде:

$$\begin{bmatrix} \hat{X}(0) \\ \hat{X}(1) \\ \vdots \\ \hat{X}(N-1) \end{bmatrix} = \sum_{n=0}^{N-1} \begin{bmatrix} A(z_0)^n \\ A(z_1)^n \\ \vdots \\ A(z_{N-1})^n \end{bmatrix} \cdot x(n). \quad (5.25)$$

Каждый элемент в данной сумме относится только к одному входному отсчету. Он может быть рассчитан при поступлении отсчета и сохранен для получения результата после поступления N отсчетов, при этом вычислительная нагрузка на входной отсчет равна $O(N)$.

5.6 Аппроксимация шкалы барков

Для достижения высокого качества реконструированных сигналов в перцептуальных системах обработки речи требуется эффективная психоакустическая модель. В известной работе Джонстона [14] психоакустическая модель основана на ДПФ: расчет ДПФ взвешенного временным окном сегмента сигнала, группировка коэффициентов преобразования в группы, соответствующие критическим частотным полосам, и расчет энергии в данных частотных полосах. Достижение приемлемого спектрального разрешения в критических частотных полосах, расположенных в низкочастотной части частотного диапазона, требует использования ДПФ с достаточно длинным временным окном. Поэтому концептуальная простота и эффективность нивелируются недостаточным временным разрешением, неприемлемым для анализа более тонкого феномена, такого как маскирование назад («pre-masking») [1].

Первый шаг при использовании WDFT в психоакустической модели – проектирование соответствующего фазового преобразования. Частотные коэффициенты z -преобразования должны быть распределены равномерно в перцептуальной шкале. В работе [15] было показано, что фазовое звено первого порядка достаточно хорошо аппроксимирует перцептуальную шкалу барков, при этом значение коэффициента фазового фильтра для заданной частоты дискретизации определяется по следующему выражению:

$$a_{Barck} = 0.1957 - 1.048 \cdot \left[\frac{2}{\pi} \arctan \left(0.07212 \cdot \frac{f_s}{1000} \right) \right]^{\frac{1}{2}}. \quad (5.26)$$

Для частоты дискретизации $f_s = 16$ кГц коэффициент $a_{Barck} = -0.57827$.

Так как ширина критических частотных полос строго изменяется с их местоположением на частотной шкале, то различное количество коэффициентов преобразования ассоциируется с конкретной критической частотной полосой. В части А табл. 5.1 количество коэффициентов в группах варьируется от 3 до 38 для ДПФ, в то время как для WDFT той же размерности не отда-

ется предпочтения ни одной из полос, все коэффициенты преобразования распределены практически равномерно (часть Б табл. 5.1).

Табл. 5.1 - Сравнение распределения коэффициентов ДПФ и WDFT в критических частотных полосах

Крит. полоса	Часть А (размер ДПФ = 512, $f_s = 16$ кГц)			Часть Б (размер WDFT = 512, $f_s = 16$ кГц)		
	Диапазон коэффициентов	Кол-во	Диапазон частот, Гц	Диапазон коэффициентов	Кол-во	Диапазон частот, Гц
1	1 – 3	3	31 - 94	1 - 12	12	8 – 100
2	4 – 6	3	125 - 188	13 - 24	12	109 – 202
3	7 – 9	3	219 - 281	25 - 36	12	210 – 305
4	10 – 13	4	313 - 406	37 - 48	12	314 – 412
5	14 – 16	3	438 - 500	49 - 60	12	421 – 523
6	17 – 20	4	531 - 625	61 - 73	13	533 – 650
7	21 – 24	4	656 - 750	74 - 85	12	660 – 776
8	25 – 29	5	781 - 906	86 - 97	12	787 – 912
9	30 – 34	5	938 - 1063	98 - 110	13	923 – 1073
10	35 – 40	6	1094 - 1250	111 - 123	13	1086 - 1254
11	41 – 46	6	1281 - 1438	124 - 135	12	1269 - 1443
12	47 – 54	8	1469 - 1688	136 - 148	13	1460 - 1680
13	55 – 62	8	1719 - 1938	149 - 161	13	1700 - 1961
14	63 – 73	11	1969 - 2281	162 - 174	13	1985 - 2302
15	74 – 86	13	2313 - 2688	175 - 186	12	2331 - 2690
16	87 – 102	16	2719 - 3188	187 - 198	12	2726 - 3174
17	103 – 122	20	3219 - 3813	199 - 210	12	3220 - 3792
18	123 – 145	23	3844 - 4531	211 - 221	11	3851 - 4513
19	146 – 173	28	4563 - 5406	222 - 231	10	4588 - 5328
20	174 – 205	32	5438 - 6406	232 - 242	11	5419 - 6412
21	206 – 243	38	6438 - 7594	243 - 252	10	6520 - 7533
22	244 – 256	13	7625 - 8000	253 - 256	4	7650 - 8000

WDFT в его оригинальной форме не сохраняет энергию сигнала в соответствующих частях единичной окружности до и после фазового звена. Вследствие этого, каждый коэффициент WDFT должен масштабироваться в соответствии с коэффициентом $\sqrt{1 - a^2}/(1 - az)$ [16], чтобы вычислить корректно уровни энергии в критических частотных полосах.

Выбор WDFT малого формата с успехом может заменить ДПФ с большой длиной выборки благодаря тому, что коэффициенты WDFT преобразования равномерно распределены в критических частотных полосах, поэтому в психоакустической модели на базе WDFT могут быть уравновешены как хорошее частотное, так и временное разрешение.

5.7 Комментарии о частотном разрешении WDFT

Представленный метод формирования избыточного базиса WDFT обнаруживает тот факт, что это равносильно результату хорошо известной процедуры добавления нулей в

«хвост» входного вектора отсчетов (от англ. zero-padding procedure). Положим, что \mathbf{D}_M есть WDFТ $M \times M$ матрица преобразования и входной вектор длины N дополнен нулями, тогда

$$\begin{bmatrix} \hat{X}(0) \\ \hat{X}(1) \\ \vdots \\ \hat{X}(M-1) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & \hat{z}_0^{-1} & \dots & \hat{z}_0^{-M+1} \\ 1 & \hat{z}_1^{-1} & \dots & \hat{z}_1^{-M+1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \hat{z}_{M-1}^{-1} & \dots & \hat{z}_{M-1}^{-M+1} \end{bmatrix}}_{\mathbf{D}_M} \cdot \begin{bmatrix} x(0) \\ \vdots \\ x(N-1) \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (5.27)$$

Заметим, что компоненты базовых векторов после \hat{z}_0^{-N+1} умножаются на добавленные нули во входной вектор и не будут влиять на конечный результат преобразования. Таким образом, выражение (5.27) эквивалентно (5.19).

Хотя число отсчетов в WDFТ с избыточным базисом увеличено, частотное разрешение в области низких частот не изменилось. Это основной вывод процедуры добавления нулей во входной вектор, показывающий, что частотное разрешение не меняется. Это только результат лучшей аппроксимации ДПФ анализируемой последовательности отсчетов. Однако, WDFТ с избыточным базисом позволяет ослабить ошибки в задаче синтеза, т.е. вычисления обратного WDFТ, но при этом, не улучшая точность преобразования. С данной точки зрения важно отметить, что даже стандартный WDFТ может быть интересен как специальный случай аппроксимации ДПФ. На рис. 5.10 и рис. 5.11 показаны примеры вычисления СПМ речевого сигнала на основе DTFT и WDFТ для $\alpha = 0$ и для $\alpha = -0.17$. Как видно из рисунков WDFТ идеально интерполирует DTFT на окне преобразования.

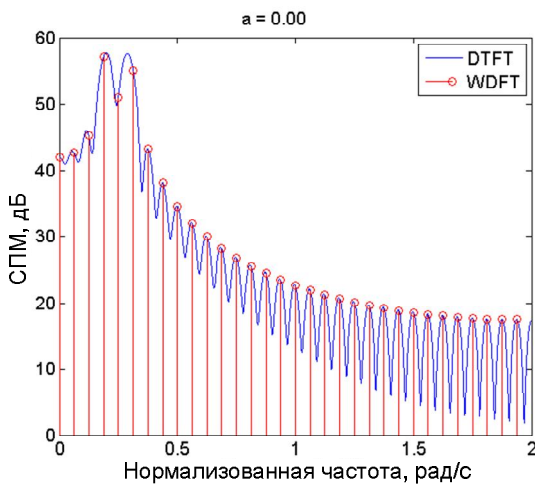


Рис. 5.10 – Вычисленная СПМ речевого сигнала на основе DTFT и WDFТ для $\alpha = 0$

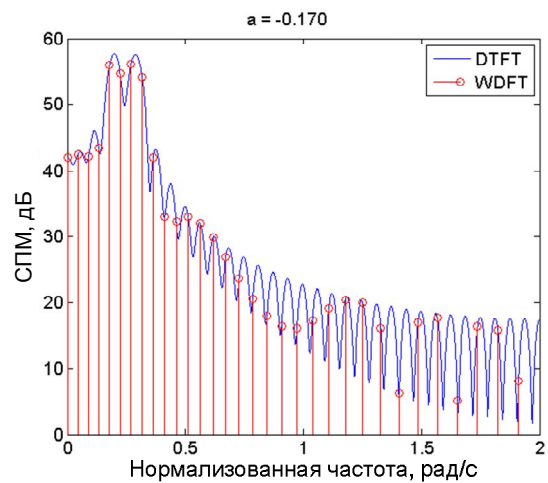


Рис. 5.11 – Вычисленная СПМ речевого сигнала на основе DTFT и WDFТ для $\alpha = -0.17$

Следовательно, фазовое преобразование не вносит никакого увеличения в частотное разрешение WDFТ. Частотное разрешение преобразования зависит только от длины и вида временного окна. Однако, различное положение частотных отсчетов может показать некоторые детали в спектре сигнала. Другими словами, точность спектрального анализа может быть повышена, но данное повышение теоретически ограничено размером окна анализа [17].

5.8 СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Zwicker E., Fastl H., Psychoacoustics: facts and models, Berlin: Springer-Verlag, 1990. 354 p.
2. Ковалгин Ю.А., Вологдин Э.И., Цифровое кодирование звуковых сигналов, СПб.: КОРОНА-принт, 2004, 240с.
3. Speech Coding and Synthesis, edited by W.B. Kleijn, K.K. Paliwal. – Elsevier, Netherlands, 1998, 743 p.
4. Martin R., and et al., A Noise Reduction Preprocessor for Mobile Voice Communication, EURASIP Journal on Applied Signal Processing, August, 2004, pp. 1046–1058.
5. Oppenheim A.V., Johnson D.H., Steiglitz K., Computation of spectra with unequal resolution using the FFT, Proceedings IEEE, February 1971, vol. 59, pp. 299–301.
6. Bagchi S., Mitra S.K., The Nonuniform Discrete Fourier Transform and Its Applications in Filter Design, IEEE Trans. Circuits Systems II, 1996, vol. 43, no. 6, pp. 422–433.
7. Nam Cho Ik., Mitra S.K., Warped discrete cosine transform and its application in image compression, IEEE Trans. Circ. Systems Video Technol, 2000, vol.10, no. 8, pp.1364-1373.
8. Makur A., Mitra S.K., A Warped Discrete-Fourier Transform: Theory and Applications, IEEE Trans. Circuits Systems I, 2001, vol. 48, no. 9, P. 1086–1093.
9. Форсайт Дж., Малькольм М., Моулер К., Машинные методы математических вычислений, М.: Мир, 1980, 279с.
10. Hansen P.C., The truncated *SVD* as a method for regularization», BIT. 1987. vol. 27. pp. 534–553.
11. Borowicz A., Parfieniuk M., Petrovsky A.A., An application of the warped discrete Fourier transform in the perceptual speech enhancement, Speech Communication, vol. 48, ELSEVIER, 2006, pp.1024-1036.
12. Penrose R., A generalized inverse for matrices, Proc. Camb. Philos. Soc., 1955, no. 51, pp. 406-413.
13. Rao C.R., Mitra S.K., Generalized inverse pf matrices and its applications, NY: Wiley, 1971.
14. Johnston J.D., Transform coding of audio signals using perceptual noise criteria, IEEE Transactions on Selected Areas Communication, February, 1988, vol. 6, pp. 314–323.
15. Smith J.O., Abel J.S., Bark and ERB Bilinear Transforms, IEEE Transactions on Speech and Audio Processing, June 1999, vol. 7, pp. 697–708.
16. Parfieniuk M., Petrovsky A., Warped DFT as the basis for psychoacoustical model, The proc. of the IEEE ICASSP, Montreal, 2004, pp. IV-185-188.
17. Петровский А.А., Борович А., Парфенюк М. Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением, Речевые технологии, №3, Москва, 2008. – С.3-15.

6 МГНОВЕННЫЙ ГАРМОНИЧЕСКИЙ АНАЛИЗ ЗВУКОВЫХ И РЕЧЕВЫХ СИГНАЛОВ

Азаров И.С., Петровский А.А.

АННОТАЦИЯ

Раздел содержит описание классических и оригинальных методов, позволяющих выполнять мгновенный гармонический анализ звуковых и речевых сигналов. Рассмотрен способ оценки мгновенных гармонических параметров на основе специального фильтра анализа, который позволяет интегрировать масштабирующую функцию для правильной оценки нестационарных периодических компонент сигнала. Приведены результаты сравнения точности различных методов оценки гармонических параметров применительно к стационарным и нестационарным сигналам.

Ключевые слова – мгновенные гармонические параметры, периодический сигнал, шумовой сигнал, основной тон, гармоники основного тона.

СОДЕРЖАНИЕ

6.1	Введение.....	115
6.2	Мгновенные гармонические параметры и классические методы их оценки.....	116
6.3	Оценка мгновенных гармонических параметров при помощи специального фильтра анализа.....	121
6.3.1	Синтез стационарного фильтра анализа.....	121
6.3.2	Синтез частотно-модулированного фильтра анализа.....	126
6.4	Сравнение точности оценки гармонических параметров различными методами	128
6.5	Список использованных источников.....	132

6.1 Введение

Изучение методов гармонического анализа является одним из важнейших направлений в современной цифровой обработке сигналов. Несмотря на имеющиеся научные достижения в этой области, вопросы спектрального и гармонического анализа до сих пор являются сложными и в значительной степени не решенными для широкого круга задач.

Гармонический анализ – это мощное средство для выполнения временно-частотных преобразований и служит одним из основных инструментов при создании современных систем мультимедиа, обрабатывающих и передающих звуковые сигналы. Под гармоническим анализом понимается способ оценки гармонических параметров для представления заданного сигнала в виде суммы периодических функций (синусоид) с меняющейся во времени амплитудой и частотой. Такой способ представления звуковой информации используется сегодня для

большого класса сигналов, в том числе звуков музыкальных инструментов и вокализованной речи.

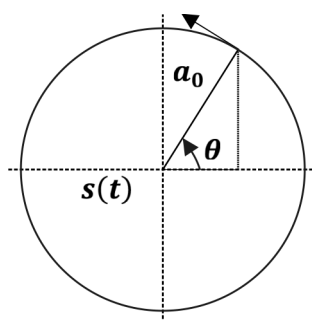
Ключевым вопросом при использовании синусоидального представления является определение гармонических параметров, так как от точности их оценки зависит результат применения всего подхода в целом. Несмотря на то, что существует большое количество способов для выполнения гармонического анализа, это научное направление постоянно совершенствуется, так как требуемая во многих приложениях точность анализа пока не достигнута. Сегодня одним из основных подходов для обработки аудио и речевой информации является комбинированное применение различных моделей. Обрабатываемый сигнал разделяется на составляющие различного свойства, после чего для обработки каждой из них применяется соответствующая модель. Этот подход в частности применяется современными параметрическими звуковыми кодерами. Одной из задач гармонического анализа является сепарация сигнала на периодическую и непериодическую компоненты. Если сепарация выполняется не точно, это заметно ухудшает работу кодера и приводит к появлению звуковых артефактов. Повышением точности разделения можно улучшить качество закодированного сигнала без каких-либо изменений в самой системе кодирования.

Для оценки мгновенных параметров существует два основных подхода: использование аналитического сигнала [1-4] и использование оператора энергии [3,5]. Оба подхода требуют выполнения предварительной фильтрации анализируемого сигнала для того, чтобы разделить его на однокомпонентные периодические функции. Однако при фильтрации возникают определенные сложности. Синтез фильтра, необходимого для выделения отдельного гармонического компонента, требует знания примерного контура его частоты, что сопряжено с использованием дополнительных передфильтрационных алгоритмов. Кроме того, возникает проблема выбора ширины полосы пропускания. Если она недостаточно узкая – фильтр будет пропускать лишние спектральные составляющие, что существенно повлияет на точность оценки. С другой стороны, если полоса недостаточно широкая, то становится невозможной точная оценка частотно-модулированных (ЧМ) периодических компонент. Поэтому для гармонического анализа сигналов, имеющих ЧМ основной тон, были предложены специальные частотно-временные преобразования, позволяющие производить адекватную оценку компонент с сильной частотной модуляцией, такие как Фан-Чирп [6] и гармоническое [7,8] преобразования. Так как они представляют собой дискретное преобразование Фурье (ДПФ) с временным масштабированием, гармонические амплитуды принимаются постоянными на протяжении всего фрейма анализа, что существенно ограничивает точность способа. Таким образом, во многих приложениях классические методы гармонического анализа не позволяют выполнять точную оценку мгновенных параметров ЧМ компонентов, и требуется применение новых, оригинальных подходов.

6.2 Мгновенные гармонические параметры и классические методы их оценки

Понятие мгновенной частоты связано с механикой и может трактоваться через движение по круговой траектории. Если тело движется с постоянной скоростью по кругу, находяще-

муся в центре координат, то проекции его движения на оси будут простыми гармоническими колебаниями (рис. 6.1).



$$s(t) = a_0 \cos \theta(t); \theta(t) = \omega_0 t$$

Рис. 6.1 – Простое гармоническое колебание $s(t)$

В момент времени t проекция $s(t)$ характеризуется смещением, постоянной скоростью и ускорением, которые определяются формулами (6.1), (6.2) и (6.3) соответственно:

$$s(t) = a_0 \cos \theta = a_0 \cos \omega t, \quad (6.1)$$

$$s'(t) = -a_0 \omega \sin \omega t, \quad (6.2)$$

$$s''(t) = a_0 \omega^2 \cos \omega t = -\omega^2 s(t). \quad (6.3)$$

Частоту $f = \omega/2\pi$ можно соотнести со смещением, решив дифференциальное уравнение (6.3). Решением является $z(t)$, определяемое как

$$z(t) = \alpha e^{j2\pi f t}, \quad (6.4)$$

где $\omega = 2\pi f$ – постоянная угловая скорость; α – произвольная константа.

Понятие мгновенной частоты было введено в 1949 году [1] на основе выражения простого гармонического колебания

$$s(t) = a \cos (2\pi f t + \theta), \quad (6.5)$$

где a – амплитуда; f – частота колебаний; θ – фазовая константа. Аргумент функции косинуса $(2\pi f t + \theta)$ является фазой $\varphi(t)$.

Амплитудные модуляции (АМ) были определены изменениями a как функции от t :

$$a(t) = a_0 [1 + \mu g(t)], \quad (6.6)$$

где $g(t)$ – модулирующий сигнал (рис. 6.2).

Аналогично фазовые модуляции были определены как

$$\theta(t) = \theta_0 [1 + \mu g(t)]. \quad (6.7)$$

Таким образом, фаза, которая является аргументом косинуса в (6.5), принимает вид: $\varphi(t) = 2\pi f t + \theta(t)$.

Для того чтобы получить выражение для сигнала с частотной модуляцией, которая определялась соответствующим образом, как функция мгновенной частоты от времени

$$f_i(t) = f_0 [1 + \mu g(t)], \quad (6.8)$$

выражение (6.5) было представлено в следующем виде:

$$s(t) = a \cos \left[\int_0^t 2\pi f_i(t) dt + \theta \right], \quad (6.9)$$

где весь аргумент косинуса представляет собой фазу $\varphi(t)$.

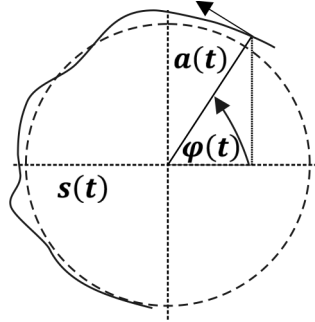


Рис. 6.2 – Амплитудная огибающая и мгновенная фаза ЧМ-АМ сигнала (сигнал определен как $s(t) = a(t)\cos\varphi(t)$; $a(t)$ – амплитудная огибающая; $\varphi(t)$ – фаза)

Запись сигнала $s(t)$ в таком виде позволяет дать мгновенной частоте следующее определение:

$$f_i(t) = \frac{1}{2\pi} \frac{d\varphi(t)}{dt}. \quad (6.10)$$

Для оценки мгновенных гармонических параметров часто используют понятие аналитического сигнала. В работе [2] предложен способ формирования уникального комплексного сигнала $z(t)$ для заданного действительного сигнала $s(t)$. Он заключается в вычислении преобразования Фурье (ПФ) действительного сигнала с последующим приравнением нулю амплитуд, принадлежащих отрицательным частотам, и удваиванием амплитуд положительных частот. Операцию можно выполнить и без перехода в частотную область:

$$z(t) = s(t) + jH[s(t)] = a(t)e^{j\varphi(t)}, \quad (6.11)$$

где H – преобразование Гильберта, определенное как

$$H[s(t)] = p. v. \int_{-\infty}^{+\infty} \frac{s(t - \tau)}{\pi\tau} d\tau, \quad (6.12)$$

где $p. v.$ обозначает главное значение интеграла по Коши.

Сигнал $z(t)$ называют аналитическим сигналом или комплексным сигналом Габора. $s(t)$ и $H[s(t)]$ теоретически отличаются по фазе на $\pi/2$, поэтому мгновенную амплитуду $a(t)$ и мгновенную фазу $\varphi(t)$ можно вычислить следующим образом:

$$a(t) = \sqrt{s^2(t) + H^2[s(t)]}, \quad (6.13)$$

$$\varphi(t) = \arctan \left(\frac{H[s(t)]}{s(t)} \right). \quad (6.14)$$

Используя средние меры Габора, в работе [9] было показано, что средняя частота спектра сигнала равна его средней мгновенной частоте:

$$\langle f \rangle = \langle f_i \rangle, \quad (6.15)$$

где

$$\langle f \rangle = \frac{\int_{-\infty}^{+\infty} f |Z(f)|^2 df}{\int_{-\infty}^{+\infty} |Z(f)|^2 df}, \quad (6.16)$$

$$\langle f_i \rangle = \frac{\int_{-\infty}^{+\infty} f_i(t) |z(t)|^2 dt}{\int_{-\infty}^{+\infty} |z(t)|^2 dt}, \quad (6.17)$$

$Z(f)$ – преобразование Фурье сигнала $z(t)$.

Аналитический сигнал для действительного дискретного сигнала имеет ту же интерпретацию и генерируется обычно при помощи ДПФ.

Кроме методов оценки, основанных на использовании аналитического сигнала, существует альтернативный подход с использованием оператора энергии [5], который для дискретного сигнала $s(n)$ вычисляется как

$$\Psi[s(n)] = s^2(n) - s(n-1)s(n+1), \quad (6.18)$$

где дифференцирование выполняется численно, как симметрическая разность.

В работе [5] предложен дискретный алгоритм сепарации энергии (DESA), который оценивает мгновенную амплитуду $MAG(n)$ и мгновенную частоту $f(n)$ сигнала $s(n)$ при помощи следующих выражений:

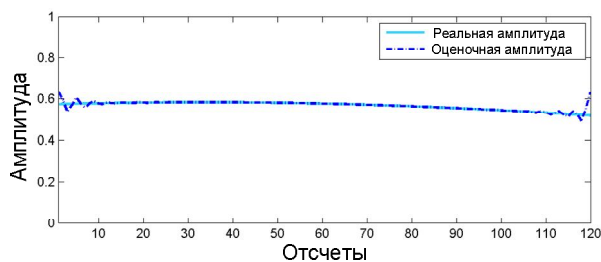
$$MAG(n) = \frac{2\Psi[s(n)]}{\sqrt{\Psi[s(n+1) - s(n-1)]}}; \quad (6.19)$$

$$f(n) = \arcsin \sqrt{\frac{\Psi[s(n+1) - s(n-1)]}{4\Psi[s(n)]}}. \quad (6.20)$$

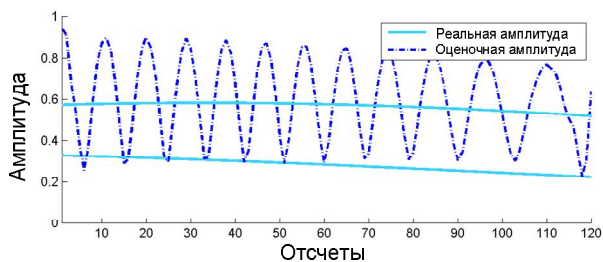
Алгоритм оценки требует всего несколько отсчетов анализируемого сигнала для вычисления гармонических параметров, что делает его весьма эффективным в использовании.

Вышеизложенные способы оценки мгновенных гармонических параметров применимы только для однокомпонентных периодических сигналов, так как однозначная функция мгновенной частоты теряет смысл тогда, когда сигнал состоит из нескольких компонентов (рис. 6.3). Поэтому для использования описанных способов в системах обработки цифрового звука предварительно применяют узкополосную фильтрацию [3,5,10], позволяющую разделить исходный сигнал на компоненты с ограниченным спектром.

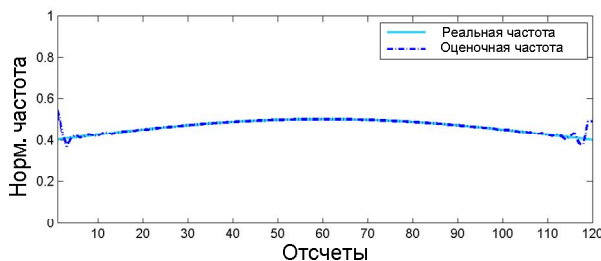
Выбор фильтров является сложной задачей, так как во многих случаях он зависит от приложения. Однако очевидно, что для применения мгновенного гармонического анализа речи и звука должна быть обеспечена фильтрация в произвольной полосе частот, что затрудняет использование готовых банков фильтров.



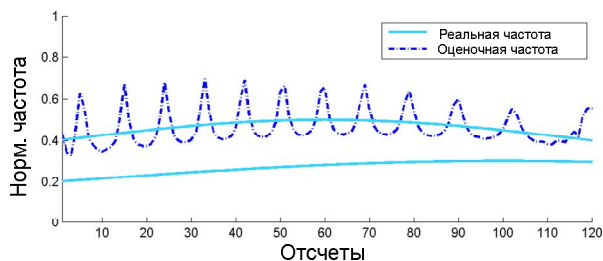
а)



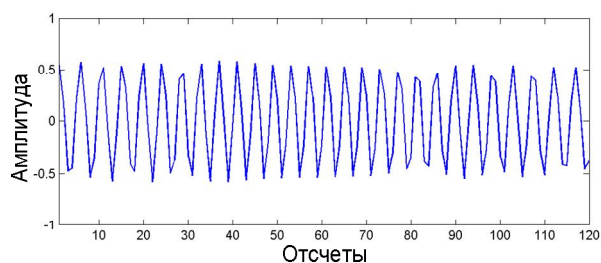
г)



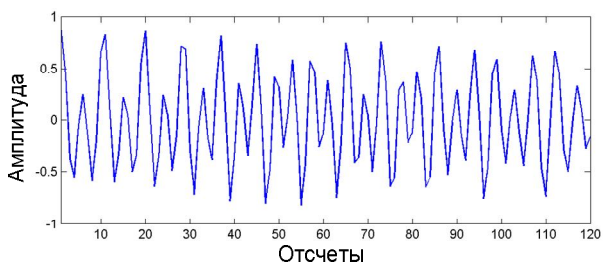
б)



д)



в)



д)

Рис. 6.3 – Анализ однокомпонентного и двухкомпонентного сигнала при помощи комплексного сигнала Габора: а) мгновенная амплитуда однокомпонентного сигнала; б) мгновенная частота однокомпонентного сигнала; в) однокомпонентный сигнал; г) мгновенные амплитуды двухкомпонентного сигнала; д) мгновенные частоты двухкомпонентного сигнала; е) двухкомпонентный сигнал

Применение методов мгновенного гармонического анализа позволяет достичь максимально высокого разрешения во временной и частотной областях одновременно, однако необходимость предварительного применения узкополосной фильтрации накладывает некоторые ограничения:

- необходимо знать примерную частоту анализируемого компонента, для того чтобы использовать нужную полосу пропускания, что требует предварительного анализа сигнала;
- фильтр должен изменять полосу пропускания от фрейма к фрейму, что в зависимости от реализации синтеза фильтра может оказаться сложным для вычисления;
- с одной стороны фильтр должен иметь как можно более высокий порядок и как можно более узкую полосу пропускания (для точного выделения анализируемого компонента), с другой – полоса пропускания должна быть максимально большой (для пропуска компоненты с частотной модуляцией).

6.3 Оценка мгновенных гармонических параметров при помощи специального фильтра анализа

Задача оценки мгновенных гармонических параметров сводится к определению амплитуды $MAG_k(n)$, частоты $f_k(n)$ и фазы $\varphi_k(n)$ для заданного дискретного момента времени из анализируемого фрейма дискретного сигнала $s(n)$, $0 \leq n < N$. Оценка параметров должна проводиться, исходя из предположения, что периодические компоненты анализируемого фрейма могут быть частотно-модулированы и их гармонические параметры меняются с каждым отсчетом.

Как было сказано выше, оценка мгновенных параметров возможна путем генерации аналитических сигналов с предварительной узкополосной фильтрацией. Однако для решения поставленной задачи требуется синтез специального фильтра анализа, имеющего следующие свойства [11]:

- фильтрация в заданной полосе частот;
- импульсная характеристика фильтра должна быть описана в виде аналитического выражения (как непрерывная функция от двух параметров – граничных частот полосы пропускания фильтра);
- оценка мгновенных параметров непосредственно из выходного сигнала (выходной сигнал может быть интерпретирован в виде аналитического сигнала);
- модулирование импульсной характеристики в соответствии с возможными частотными модуляциями анализируемого компонента (неявное временное масштабирование);
- непрерывные и гладкие траектории выходных параметров $MAG_k(n)$ и $f_k(n)$.

Таким образом, узкополосная фильтрация, с одной стороны, может обеспечить разделение анализируемого сигнала на периодические компоненты и получение мгновенных гармонических параметров, с другой – позволит правильно обрабатывать компоненты с частотной модуляцией. Данный подход объединяет в себе элементы преобразований со встроенным временным масштабированием (Фан-Чирп преобразование [6] и гармоническое преобразование [7,8]) и методов оценки на основе аналитических сигналов. Возможность определения мгновенных параметров частотно-модулированных компонент может дополнительно повысить качество гармонического анализа речевых сигналов в системах мультимедиа.

6.3.1 Синтез стационарного фильтра анализа

Фильтр анализа может быть синтезирован используя N -точечное преобразование Фурье [12], которое можно рассматривать как фильтр с конечной импульсной характеристикой, формирующий на выходе стационарную синусоиду $\bar{s}(n)$ заданной нормализованной частоты f

$$\bar{s}(n) = MAG(S(f)) \cos\left(\frac{2\pi n f}{N} + \varphi(S(f))\right), \quad (6.21)$$

чья постоянная амплитуда $MAG(S(f))$ и начальная фаза $\varphi(S(f))$ может быть вычислена следующим образом:

$$\text{MAG}(S(f)) = \sqrt{\text{Re}S(f)^2 + \text{Im}S(f)^2}; \quad (6.22)$$

$$\varphi(S(f)) = -\arctan \frac{\text{Im}S(f)}{\text{Re}S(f)}. \quad (6.23)$$

Импульсная характеристика этого фильтра $h(n)$ может быть записана в виде непрерывной аналитической функции для частоты f в Гц:

$$h(n) = \cos\left(\frac{2\pi}{F_s}nf\right). \quad (6.24)$$

Используя (6.24), можно получить выражение для импульсной характеристики фильтра, формирующего на выходе синусоидальный компонент с ограниченным частотным диапазоном

$$h(n) = \frac{\int_{F_1}^{F_2} \cos\left(\frac{2\pi}{F_s}nf\right) df}{F_2 - F_1}, \quad (6.25)$$

где F_1 и F_2 границы полосы пропускания ($F_1 < F_2$).

Интегрирование выражения (6.25) с раскрытием неопределенности при $n = 0$ ведет к импульсной характеристике фильтра в следующем виде:

$$h(n) = \begin{cases} 1, & n = 0 \\ \frac{\frac{F_s}{n\pi} \cos\left(\frac{2\pi n}{F_s}F_c\right) \sin\left(\frac{2\pi n}{F_s}F_\Delta\right)}{2F_\Delta} & n \neq 0 \end{cases}, \quad (6.26)$$

где $F_c = (F_1 + F_2)/2$, $F_\Delta = (F_2 - F_1)/2$.

Параметры F_c и F_Δ соответствуют центральной частоте полосы пропускания и половине ее ширины соответственно. Сигнал на выходе фильтра $s_{F_c, F_\Delta}(n)$ может быть вычислен как свертка $s(n)$ с $h(n)$ и записан в виде следующей суммы:

$$s_{F_c, F_\Delta}(n) = \sum_{i=0}^{N-1} \frac{s(i)F_s}{2\pi(n-i)F_\Delta} \cos\left(\frac{2\pi(n-i)}{F_s}F_c\right) \sin\left(\frac{2\pi(n-i)}{F_s}F_\Delta\right). \quad (6.27)$$

Выражение (6.27) запишем следующим образом:

$$s_{F_c, F_\Delta}(n) = A(n) \cos(0n) + B(n) \sin(0n), \quad (6.28)$$

где

$$A(n) = \sum_{i=0}^{N-1} \frac{s(i)F_s}{2\pi(n-i)F_\Delta} \sin\left(\frac{2\pi(n-i)}{F_s}F_\Delta\right) \cos\left(\frac{2\pi(n-i)}{F_s}F_c\right),$$

$$B(n) = \sum_{i=0}^{N-1} \frac{-s(i)F_s}{2\pi(n-i)F_\Delta} \sin\left(\frac{2\pi(n-i)}{F_s}F_\Delta\right) \sin\left(\frac{2\pi(n-i)}{F_s}F_c\right).$$

Учитывая (6.28), сигнал на выходе фильтра (6.27) представляет собой АМ и ЧМ функцию косинуса:

$$s_{F_c, F_\Delta}(n) = \text{MAG}(n) \cos(\varphi(n)) \quad (6.29)$$

с мгновенной амплитудой $\text{MAG}(n)$, фазой $\varphi(n)$ и частотой $f(n)$, которые могут быть определены по следующим формулам:

$$\text{MAG}(n) = \sqrt{A^2(n) + B^2(n)}, \quad (6.30)$$

$$\varphi(n) = \arctan\left(\frac{-B(n)}{A(n)}\right), \quad (6.31)$$

$$f(n) = \frac{\varphi(n+1) - \varphi(n)}{2\pi} F_s. \quad (6.32)$$

Так как фаза компонент $A(n)$ и $B(n)$ отличается на $\pi/2$, выход фильтра анализа $s_{F_c, F_\Delta}(n)$ может быть записан в виде аналитического сигнала $s_{F_c, F_\Delta}^a(n)$:

$$s_{F_c, F_\Delta}^a(n) = A(n) + jB(n). \quad (6.33)$$

Можно показать, что при разбиении всего частотного диапазона сигнала произвольными частотами F_1, F_2, \dots, F_L такими, что $F_1 = 0; F_L = F_s/2; F_1 \leq F_2 \leq \dots \leq F_L$ и $L - 1$, исходный сигнал $s(n)$ равняется сумме выходов фильтров анализа:

$$s(n) = \sum_{k=1}^{L-1} s_{F_k, F_{k+1}}(n). \quad (6.34)$$

Сложение соответствующих аналитических сигналов $s_{F_k, F_k}^a(n)$ дает в результате комплексный сигнал Габора, соответствующий исходному сигналу $s(n)$, который вычисляется при помощи соотношения (6.11):

$$z(n) = \sum_{k=1}^{L-1} s_{F_k, F_k}^a(n) = s(n) + jH[s(n)]. \quad (6.35)$$

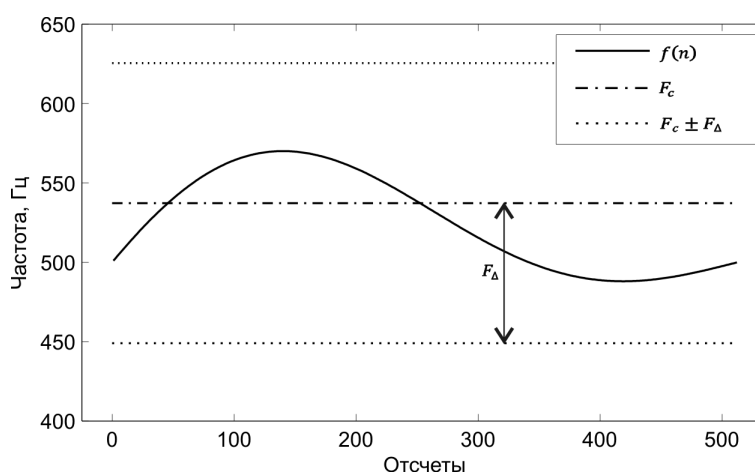


Рис. 6.4 – Кратковременный гармонический анализ с использованием стационарного фильтра

Мгновенные гармонические параметры выходного сигнала могут быть рассчитаны в любой момент времени, принадлежащий анализируемому фрейму сигнала, причем этот мо-

мент не ограничивается дискретными отсчетами сигнала, так как выход фильтра записан в виде непрерывных функций [13]. Очевидно, что полоса пропускания фильтра, задаваемая параметрами F_1 и F_2 (или F_C и F_Δ), должна содержать анализируемый компонент, и в случае наличия частотной модуляции полоса может оказаться достаточно большой (рис. 6.4).

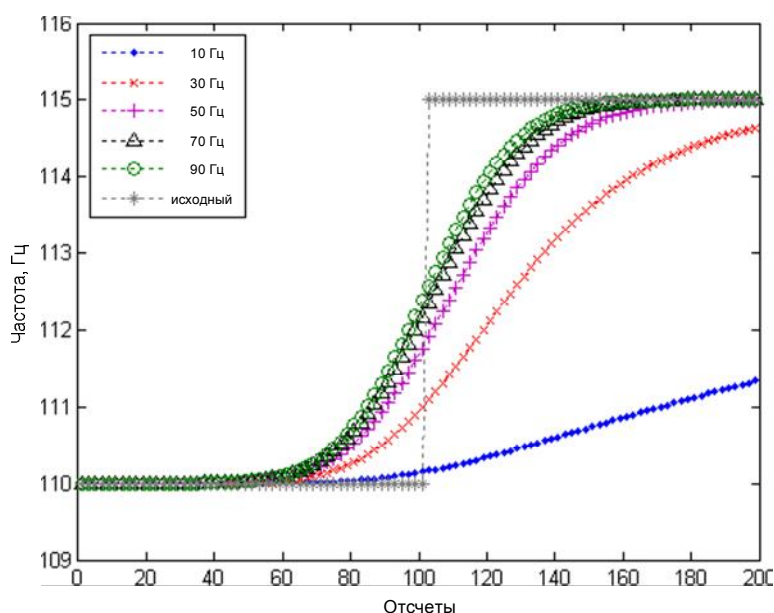


Рис. 6.5 – Оценка параметров частотно-модулированного гармонического компонента при помощи фильтров анализа с разной полосой пропускания (ширина полосы пропускания 10 – 90 Гц)

На рис. 6.5 показано, как расширение полосы пропускания фильтра влияет на точность оценки мгновенной частоты. Анализируемый гармонический компонент синтезирован с разрывом контура частоты, что дает возможность определить скорость реакции фильтров анализа.

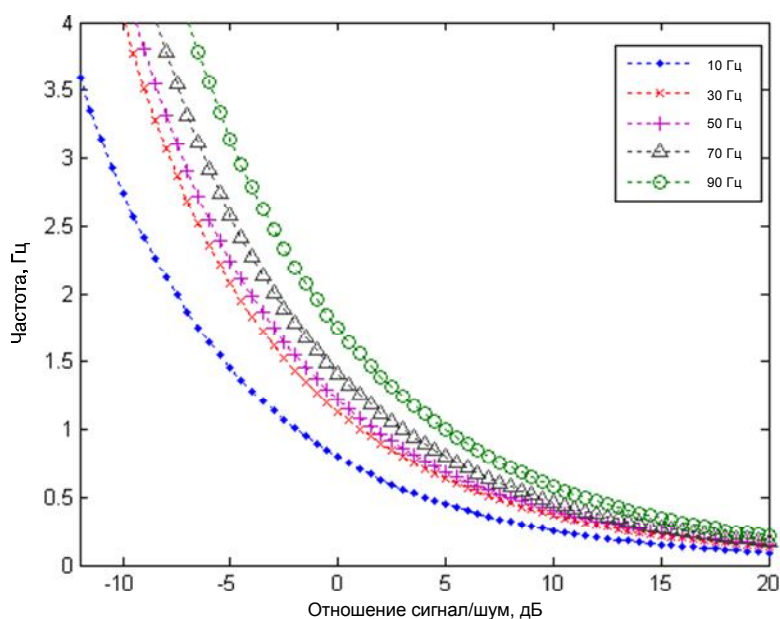


Рис. 6.6 – Средняя ошибка оценки мгновенной частоты, в зависимости от уровня белого шума в сигнале для фильтров анализа с разной полосой пропускания (ширина полосы пропускания 10 – 90 Гц)

Как было отмечено ранее, расширение полосы пропускания нежелательно, так как это приводит к попаданию дополнительных составляющих в выходной сигнал и, как следствие, к ухудшению точности анализа. На рис. 6.6 показано, как ширина полосы пропускания фильтра влияет на точность оценки мгновенной частоты, если анализируемый сигнал содержит белый шум [14]. Здесь используется отношение гармоники/шум (HNR), как мера содержания шума в сигнале:

$$HNR = 10 \lg \frac{E_h}{E_r}, \quad (6.36)$$

где E_h и E_r являются энергиями гармонической и шумовой составляющих соответственно.

В любом случае ширина полосы пропускания фильтра должна быть достаточно большой для того, чтобы выполнять адекватную оценку гармонических амплитуд. В том случае, когда полоса слишком узкая, оценочная амплитуда будет меньше реальной. Минимальная ширина полосы пропускания определяется таким образом, чтобы не занижать оцениваемый амплитудный параметр более, чем на указанный порог. Она зависит от длины окна анализа (продолжительность анализируемого фрейма) и используемой оконной функции.

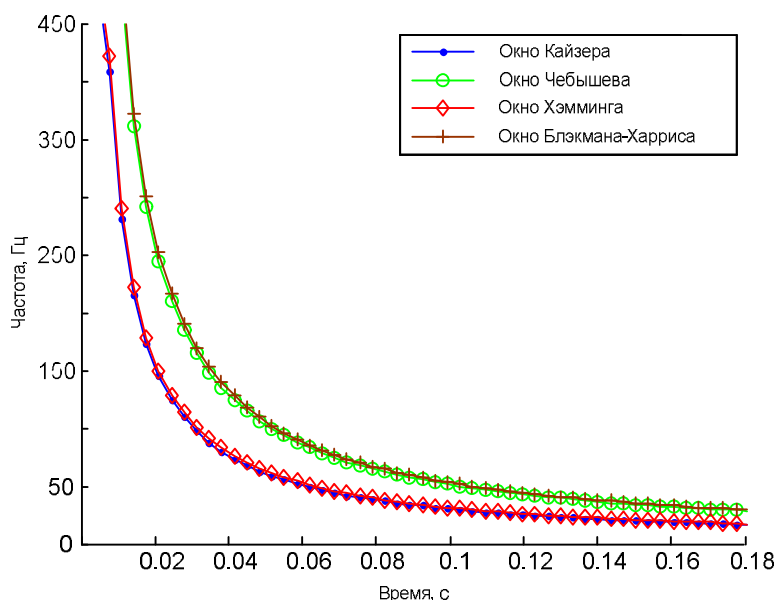


Рис. 6.7 – Зависимость минимальной ширины полосы пропускания от продолжительности анализируемого фрейма при допустимом понижении амплитуды не более чем на -20дБ

Из рис. 6.7 следует, что необходимая ширина полосы уменьшается с увеличением продолжительности анализируемого фрейма. Следовательно, для наиболее точной оценки следует использовать продолжительные фреймы, что не всегда допустимо из-за частотной модуляции гармонических компонент.

Таким образом, в отдельных приложениях (вокализованная речь) для точного анализа частотно-модулированных компонент нужно использовать специальный фильтр, адаптирующийся к изменяющемуся частотному контуру.

6.3.2 Синтез частотно-модулированного фильтра анализа

Для точной оценки синусоидальных параметров компонентов с сильной частотной модуляцией должен быть использован частотно-модулированный фильтр [15,16], импульсная характеристика которого модулируется в соответствии с частотным контуром анализируемого компонента. Данный подход хорош для анализа вокализованной речи, так как приблизительные частотные траектории гармоник могут быть получены из контура частоты основного тона. Непрерывность импульсной характеристики синтезированного стационарного фильтра позволяет достаточно просто адаптировать его к частотным модуляциям. Рассматривая центральную частоту полосы пропускания, как функцию от времени $F_c(n)$, выражение импульсной характеристики фильтра анализа (6.26) может быть представлено в следующем виде:

$$h(n) = \begin{cases} 1, & n = 0 \\ \frac{\frac{F_s}{n\pi} \cos\left(\frac{2\pi}{F_s} \varphi_c(n, i)\right) \sin\left(\frac{2\pi n}{F_s} F_\Delta\right)}{2F_\Delta} & n \neq 0 \end{cases}, \quad (6.37)$$

где

$$\varphi_c(n, i) = \begin{cases} \sum_{j=n}^i F_c(j), & n < i \\ -\sum_{j=i}^n F_c(j), & n > i \\ 0, & n = i \end{cases}.$$

Тогда выходной сигнал фильтра будет иметь следующий вид:

$$s_{F_c, F_\Delta}(n) = \sum_{i=0}^{N-1} \frac{s(i)F_s}{2\pi(n-i)F_\Delta} \cos\left(\frac{2\pi}{F_s} \varphi_c(n, i)\right) \sin\left(\frac{2\pi(n-i)}{F_s} F_\Delta\right), \quad (6.38)$$

что можно записать как

$$s_{F_c, F_\Delta}(n) = A(n) \cos(0n) + B(n) \sin(0n), \quad (6.39)$$

где

$$A(n) = \sum_{i=0}^{N-1} \frac{s(i)F_s}{2\pi(n-i)F_\Delta} \sin\left(\frac{2\pi(n-i)}{F_s} F_\Delta\right) \cos\left(\frac{2\pi}{F_s} \varphi_c(n, i)\right),$$

$$B(n) = \sum_{i=0}^{N-1} \frac{-s(i)F_s}{2\pi(n-i)F_\Delta} \sin\left(\frac{2\pi(n-i)}{F_s} F_\Delta\right) \sin\left(\frac{2\pi}{F_s} \varphi_c(n, i)\right).$$

Требуемые мгновенные параметры (амплитуда, частота, фаза), так же как и в случае стационарного фильтра, могут быть вычислены при помощи выражений (6.30) – (6.32).

Частотно-модулированный фильтр имеет масштабированную в частотной области полосу пропускания, задаваемую частотным контуром $F_c(n)$, что обеспечивает адекватный анализ

периодических компонент с быстрым изменением частоты. Этот подход является альтернативой временному масштабированию, которое применяется в анализе речи [6,10].

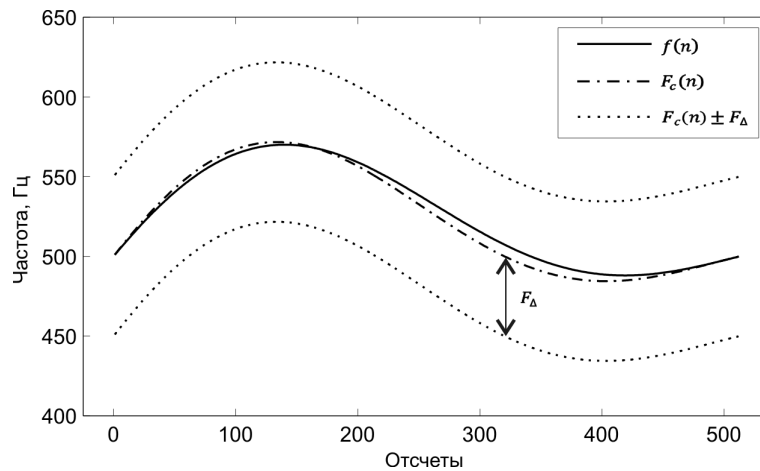


Рис. 6.8 – Кратковременный гармонический анализ с использованием частотно-модулированного фильтра

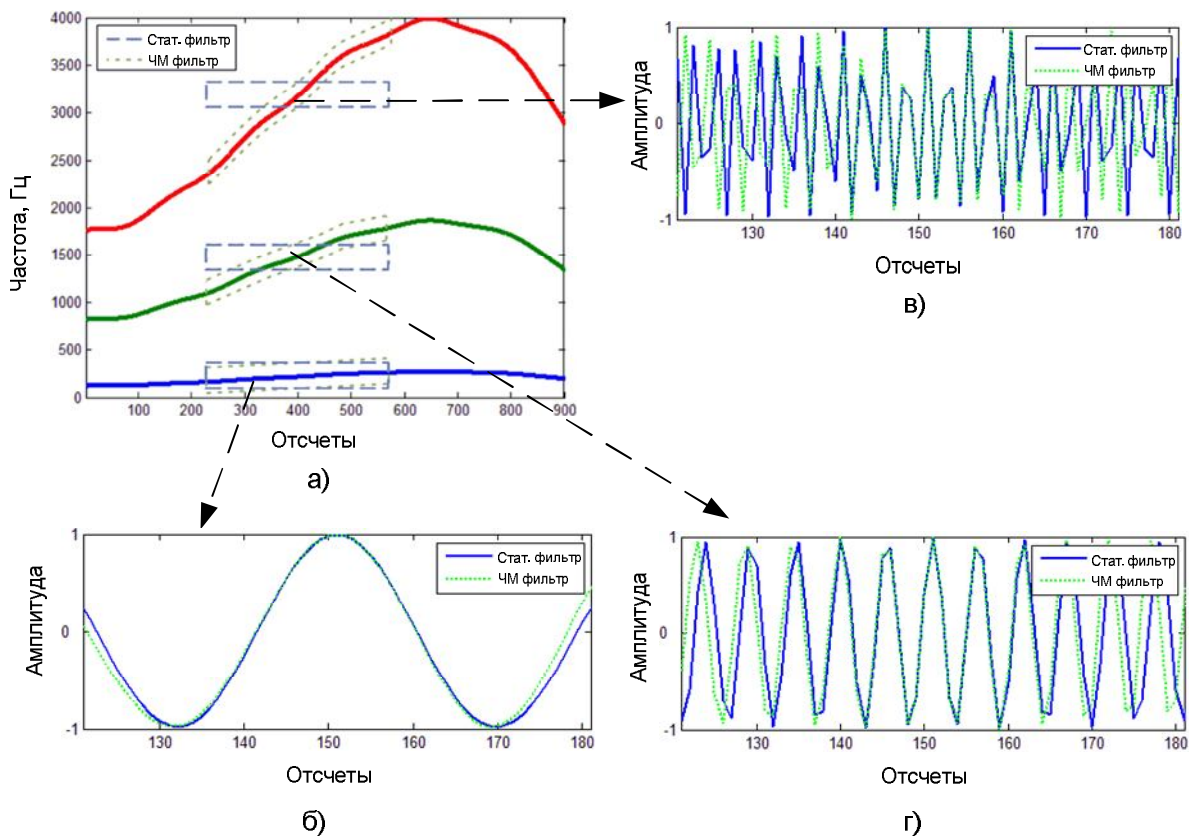


Рис. 6.9 – Применение стационарного и ЧМ фильтра для анализа вокализованной речи: а) – полосы пропускания фильтров и частотные контура гармоник; б) - г) – импульсные характеристики фильтров для оценки основного тона 7-й и 15-й гармоник соответственно

Таким образом, анализируемый частотно-модулированный компонент может быть выделен при помощи узкой масштабированной полосы пропускания (рис. 6.8). Сужение полосы обеспечивает более точную локализацию и, соответственно, делает оценки параметров более точными. Это так же позволяет применять мгновенный гармонический анализ к гармони-

кам высокого порядка речевого сигнала (рис. 6.9). Чем выше номер гармоники, тем больше изменение ее частоты, и импульсная характеристика ЧМ фильтра изменяется соответствующим образом.

6.4 Сравнение точности оценки гармонических параметров различными методами

Точность оценки гармонических параметров оценивается путем определения качества сепарации тестовых сигналов на периодическую и шумовую компоненты [17]. В качестве тестовых сигналов используются синтезированные звуки, представляющие собой комбинацию гармонической и шумовой составляющих, соотношение которых задается коэффициентом отношения гармоники/шум HNR

$$HNR = 10 \lg \frac{\sigma_H^2}{\sigma_e^2}, \quad (6.40)$$

где σ_H^2 – энергия гармонической составляющей тестового сигнала; σ_e^2 – энергия шумовой компоненты тестового сигнала.

В процессе тестирования выполнялась сепарация сигналов различными методами, и вычислялись следующие оценки.

Отношение энергии периодической составляющей, синтезированной при помощи оценочных значений гармонических параметров $\tilde{\sigma}_H^2$, к энергии шумовой компоненты σ_e^2 :

$$\tilde{HNR} = 10 \lg \frac{\tilde{\sigma}_H^2}{\sigma_e^2}. \quad (6.41)$$

Отношение энергии периодической компоненты тестового сигнала σ_H^2 к энергии синтезируемой периодической компоненты $\tilde{\sigma}_H^2$:

$$\frac{O}{E} = 10 \lg \frac{\sigma_H^2}{\tilde{\sigma}_H^2}. \quad (6.42)$$

Отношение сигнал/шум синтезируемой периодической компоненты

$$SNR_H = 10 \lg \frac{\tilde{\sigma}_H^2}{\sigma_{eH}^2}, \quad (6.43)$$

где σ_{eH}^2 – энергия сигнала, вычисляемого как разница между исходной периодической частью тестового сигнала и синтезированной при помощи оценочных значений.

Процедура оценки содержит следующие этапы:

Этап 1. Формирование тестового сигнала.

Этап 2. Гармонический анализ и синтез периодической компоненты на основании полученных значений амплитуд и фаз ее гармонических составляющих.

Этап 3. Оценка шумовой компоненты тестового сигнала.

Этап 4. Измерение оценок качества сепарации сигнала: \tilde{HNR} , O/E , SNR_H .

Этап 5. Повтор этапов (*операций*) 1-4 10 раз для каждого тестового сигнала, каждого метода и каждого выбранного соотношения HNR с целью получения средних значений.

Процедура гармонического анализа выполнялась идентично для всех сигналов. Для сравнения результатов, использовались различные способы, применяющиеся для получения частотно-временного представления сигнала. Всего использовалось пять различных методов:

Метод 1 – дискретное преобразование Фурье с частотным разрешением, адаптированным к частоте основного тона;

Метод 2 – Фан-Чирп преобразование [6];

Метод 3 – выделение гармонических составляющих при помощи узкополосной фильтрации с последующим вычислением мгновенных гармонических параметров при помощи аналитических сигналов (использовалось преобразование Гильберта);

Метод 4 – способ гармонического преобразования при помощи стационарного фильтра анализа;

Метод 5 – способ гармонического преобразования при помощи частотно-модулированного фильтра анализа.

Во всех методах использовался один и тот же заданный контур частоты основного тона, окно анализа одинаковой продолжительности (64 мс), одна оконная функция (окно Хэмминга). Для методов 3-5 использовалась одинаковая ширина полосы пропускания фильтра – 56 Гц.

Табл. 3.1 – Результат анализа сигнала 1

	Метод 1	Метод 2	Метод 3	Метод 4	Метод 5
HNR	∞				
$\tilde{H}NR$	41.5002	41.5002	26.0936	50.4436	50.4436
O/E	0.0721	0.0721	0.4203	-0.0261	-0.0261
SNR_H	41.5723	41.5723	26.5139	50.4175	50.4175
HNR	40				
$\tilde{H}NR$	38.0548	38.0548	25.9638	41.1487	41.1517
O/E	0.0677	0.0677	0.4163	-0.0299	-0.0299
SNR_H	41.3692	41.3692	26.5539	44.6628	44.6629
HNR	20				
$\tilde{H}NR$	20.7832	20.7832	20.2433	21.9468	21.9468
O/E	0.0585	0.0585	0.3952	-0.0504	-0.0504
SNR_H	29.2566	29.2566	23.7440	26.2198	26.2198
HNR	10				
$\tilde{H}NR$	10.7224	10.7224	11.4217	11.8900	11.8900
O/E	0.1582	0.1582	0.4554	0.0103	0.0103
SNR_H	19.5001	19.5001	16.1230	16.4012	16.4012
HNR	0				
$\tilde{H}NR$	1.2764	1.2764	2.5739	2.9199	2.9200
O/E	-0.2777	-0.2777	-0.4037	-0.8449	-0.8451
SNR_H	9.2242	9.2242	6.3599	5.9978	5.9980

Таким образом, результаты тестирования объективно отражают, как описанный в данной работе способ оценки мгновенных гармонических параметров на основе фильтров анализа соотносится с другими, использующимися на сегодняшний день, способами.

Ниже приведено описание тестовых сигналов и результаты сравнения вышеперечисленных методов.

Сигнал 1:

- основной тон – 150 Гц, частота основного тона постоянна;
- начальное распределение фаз гармонических компонентов случайное;
- распределение амплитуд случайное, амплитуды постоянны;
- значения частот гармонических компонентов строго кратны частоте основного тона;
- число гармоник – 20;
- шумовая компонента – аддитивный белый шум, энергия которого соответствует соотношению HNR .

Результаты анализа сигнала приведены в табл. 3.1.

Табл. 3.2 – Результат анализа сигнала 2

	Метод 1	Метод 2	Метод 3	Метод 4	Метод 5
HNR	∞				
$\tilde{H}NR$	9.2339	13.3273	19.1610	22.3343	26.9543
O/E	1.1600	0.7238	0.5939	0.0271	0.1071
SNR_H	10.3940	14.0511	19.7548	22.3614	27.0614
HNR	40				
$\tilde{H}NR$	8.1534	10.5788	17.5173	20.1762	24.8233
O/E	1.6166	1.2681	0.7438	0.1832	0.2113
SNR_H	9.8002	11.8966	18.3674	20.4839	25.3210
HNR	20				
$\tilde{H}NR$	8.3533	11.9102	16.2526	17.9287	19.3424
O/E	1.3036	0.8511	0.5932	0.0205	0.0929
SNR_H	10.0262	13.6454	18.2336	19.7145	22.6816
HNR	10				
$\tilde{H}NR$	5.9919	6.8508	8.9600	9.5400	9.6224
O/E	0.7218	0.4894	0.3692	-0.1533	-0.1165
SNR_H	10.4311	12.1108	13.6517	13.6647	14.0270
HNR	0				
$\tilde{H}NR$	-1.9943	-1.5969	0.1030	0.5188	0.4751
O/E	0.0277	-0.2517	-0.8418	-1.3492	-1.3011
SNR_H	5.1287	6.1102	4.4684	4.1168	4.2396

Сигнал 2:

- основной тон нелинейно изменяется со скоростью 0,08 Гц/мс;
- начальное распределение фаз гармонических компонентов случайное;
- распределение амплитуд последовательно моделирует несколько гласных звуков;
- значения частот гармонических компонентов отклоняются от значений кратных частоте основного тона случайным образом на ± 10 Гц;

- число гармоник – 20;
- шумовая компонента – аддитивный белый шум, энергия которого соответствует соотношению HNR .

Результаты анализа сигнала приведены в табл. 3.2.

На рис. 6.10 показаны значения SNR_H , полученные для различных методов.

Как следует из представленных экспериментальных результатов, методы 3-5 более чувствительны к уровню шума, чем методы 1-2, потому что они анализируют компоненты сигнала в более широкой частотной полосе. Тем не менее, при малом уровне аддитивного шума ($40\text{dB} < HNR < \infty$) методы 4-5 превосходят методы 1-2 для обоих тестовых сигналов. Наличие широкой полосы анализа так же является преимуществом при работе с ЧМ и АМ сигналами. Поэтому в случае, наиболее близкому к натуральному гармоническому сигналу (вокализованная речь – сигнал 2), методы 1-2 значительно уступают методу 5 вплоть до $HNR = 10$ дБ. Из результатов эксперимента можно сделать вывод о том, что в речевых приложениях метод 5 является наиболее предпочтительным для гармонического анализа, так как он позволяет выполнять наиболее точное разделение периодической и шумовой составляющих нестационарного сигнала. Исключение составляет сильно зашумленный сигнал ($HNR < 10$ дБ), однако на практике сигнал с таким уровнем шума обычно требует применения специальных методов шумоподавления перед выполнением гармонического анализа вне зависимости от применяемого способа частотно-временного преобразования.

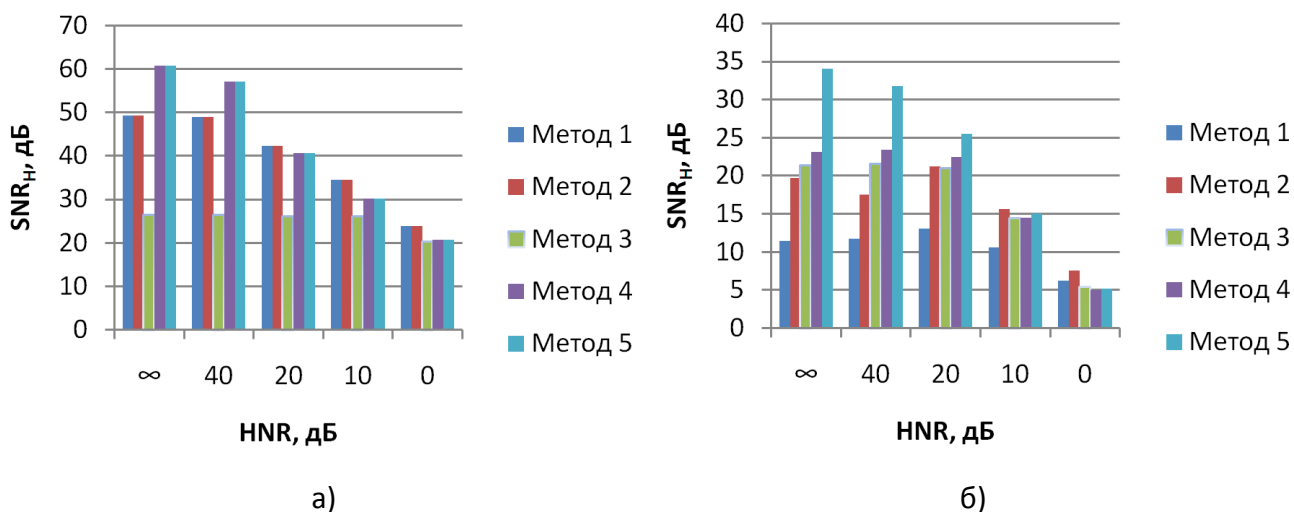


Рис. 6.10 – Отношения SNR_H , полученные для разных сигналов и разных методов оценки:
а – сигнал 1; б – сигнал 2

Концепция использования мгновенных гармонических параметров для частотно-модулированных сигналов с целью получения более точного параметрического описания звука и речи позволяет усовершенствовать системы мультимедиа для обработки звука (системы синтеза речи по тексту, кодеры и вокодеры, конверторы голоса и голосовые процессоры). Приведенные результаты свидетельствуют о перспективности дальнейших исследований в области гармонического анализа, а так же методов обработки гармонических параметров.

Перспектива развития гармонической модели и методов гармонического преобразования заключается в поиске специфических способов анализа и обработки, зависящих от целе-

вого приложения системы. Одним из достоинств предлагаемого подхода является гибкость и адаптивность к требованиям решаемых задач. Преимуществом так же является гибкость представления акустической информации, которая адаптируется необходимым образом, отвечая требованиям той или иной задачи.

6.5 Список использованных источников

1. Boashash B, Estimating and interpreting the instantaneous frequency of a signal, Proceedings of the IEEE, 1992, Vol.80, № 4, P. 520-568.
2. Gabor D., Theory of communication, Proc. IEE., 1946. Vol.93, no. 3, pp. 429-457.
3. Gianfelici F. and et al., Multicomponent AM–FM Representations: An Asymptotically Exact Approach, IEEE Transactions on Audio, Speech, and Language Processing, 2007, vol.15, no 3, pp. 823-837.
4. Hahn S.L., Hilbert transforms in signal processing, Boston: Artech House Publishers, 1996. 460 p.
5. Maragos P., Kaiser J. F., Quatieri T. F., Energy separation in signal modulations with application to speech analysis, IEEE Transactions on Signal Processing, 1993, vol.41, no 10, pp. 3024-3051.
6. Weruaga L., Kepesi M., The fan-chirp transform for non-stationary harmonic signals, Signal Processing, 2007, vol.87, no 6, pp. 1-18.
7. Zhang F., Bi G., Chen Y.Q., Harmonic transform, IEEE Proc.-Vis. Image Signal Process, 2004, vol.151, no 4, pp. 257-264.
8. Zubrycki P., Petrovsky A., Accurate speech decomposition into periodic and aperiodic components based on discrete harmonic transform, Signal processing: proceedings of 15-th europ. conf. (EUSIPCO'2007), Poznan, Poland, 2007, Poznan, 2007, pp. 2336–2340.
9. Ville J., Theorie et application de la notion de signal analytic, Cables et Transmissions, 1948, vol. 2A, no 1, pp. 61-74.
10. Abe T., Honda M., Sinusoidal model based on instantaneous frequency attractors, IEEE Trans. on Audio, Speech, and Language processing, 2006, vol. 14, no. 4, pp. 1292-1300.
11. Азаров И.С., Петровский А.А., Вычисление мгновенных гармонических параметров речевого сигнала, Речевые технологии, 2008, № 1 (1), с. 67–77.
12. Azarov E. Petrovsky A., Estimation of the instantaneous harmonic parameters of speech, Recent advances in signal processing, Vienna: In-Tech publisher, 2009, 23 pages.
13. Azarov E., Petrovsky A., Instantaneous harmonic analysis for vocal processing [electronic resource], DAFx-09: proc. of the 12th International Conference on Digital Audio Effects, Italy, Como, September 1-4, 2009, Como., 2009, Mode of access: http://dafx09.como.polimi.it/proceedings/papers/paper_25.pdf, Date of access: 04.09.2009.

-
14. Azarov E., Petrovsky A., Parfieniuk M., Estimation of the instantaneous harmonic parameters of speech [electronic resource], EUSIPCO'08: proc. of the 16th European Signal Processing Conference, Switzerland, Lausanne, August 25-29, 2008, Mode of access: <http://eurasip.org/Proceedings/Eusipco/Eusipco2008/papers/1569101820.pdf>, Date of access: 09.09.2008.
 15. Petrovsky A., Stankevich A., Balunowski J., The order tracking front-end algorithms in the rotating machine monitoring systems based on the new digital low order tracking, On sound and vibration: Proceedings of the 6th Intern. congress, Copenhagen, Denmark, 1999, pp. 2985-2992.
 16. Petrovsky A.A., The synthesis of high order digital band-pass filters with tunable center frequency and bandwidth, Signal processing: proceedings of 8th European Signal Processing Conference (EUSIPCO'96), Trieste, Italy, September 10–13, 1996, pp. 1527–1530.
 17. Yegnanarayana B, d'Alessandro C., Darsinos V., An iterative algorithm for decomposition of speech signals into periodic and aperiodic components, IEEE Trans. On Speech and Audio Proc., 1998, vol. 6, no. 1, pp. 1-11.

7 ПСИХОАКУСТИЧЕСКИЕ МОДЕЛИ ДЛЯ СИСТЕМ МУЛЬТИМЕДИА НА ОСНОВЕ СУБПОЛОСНОЙ ОБРАБОТКИ СИГНАЛОВ

Петровский Ал.А., Парфенюк М.

АННОТАЦИЯ

Цель данной главы – показать методологию построения психоакустической модели для трех способов субполосной обработки речевых и звуковых сигналов на основе пакета дискретного вейвлет преобразования, многополосного вейвлет преобразования с использованием гиперкомплексной арифметики – кватернионов, ДПФ с неравномерным частотным разрешением (от англ. Warped discrete Fourier transform (WDFT)), когда ошибка аппроксимации критических частотных полос минимальна в области Барков, а также получить процедуру расчета маскирующих порогов в пространстве коэффициентов соответствующего преобразования. Субъективные оценки качества восстановленной речи в системах с перцептуальной обработкой сигналов на данных подходах выше. Это объясняется тем, что весь процесс обработки осуществляется в перцептуальном домене с неравномерной частотной шкалой и нет необходимости в преобразованиях между разными частотными шкалами, что приводит к упрощению архитектуры системы.

Ключевые слова: порог маскирования, перцептуальная обработка, ПДВП, параунитарный банк фильтров, WDFT

СОДЕРЖАНИЕ

7.1	Введение.....	135
7.2	ПДВП, согласованное со шкалой критических частотных полос.....	136
7.3	Процедура расчета порогов маскирования в вейвлет области.....	142
7.4	Перцептуальная обработка в домене 4-х полосного параунитарного банка фильтров на основе кватернионов.....	147
7.4.1	Многополосные вейвлеты	147
7.4.2	Решетчатая структура М-канального параунитарного банка фильтров с линейной фазой на кватернионах	148
7.4.3	Психоакустическая модель на основе 4-х полосного вейвлет преобразования на кватернионах	151
7.5	Психоакустическая модель на основе WDFT ..	153
7.5.1	Основные принципы WDFT.....	153
7.5.2	Аппроксимация шкалы барков	154
7.5.3	Оценка порогов маскирования в WDFT-области.....	157
7.5.4	Эмпирическое сравнение моделей аппроксимации критических частотных полос	159
7.6	Список использованных источников.....	161

7.1 Введение

Субполосная перцептуальная обработка сигналов звука и речи [1-3] осуществляется в критических частотных полосах согласно психоакустической модели Zwicker [4]:

- расстояние между центральными частотами соседних критических частотных полос

$$z = F(f) = 13 \cdot \arctan(0.00076 \cdot f) + 3.5 \cdot \arctan\left(\left(\frac{f}{7500}\right)^2\right), [\text{Барк}] \quad (7.1)$$

где f – частота в герцах, единица измерения в данном масштабе - 1 Барк;

- ширина критических частотных

$$CBW(f) = 25 + 75 \cdot \left(1 + 1.4 \cdot \left(\frac{f}{1000}\right)^2\right)^{0.69} [\text{Гц}]. \quad (7.2)$$

При этом, также вычисляются характеристики психоакустической модели восприятия человеком акустической информации, такие как пороги маскирования:

- абсолютный порог слышимости ATH (absolute threshold of hearing), частотная зависимость которого аппроксимируется выражением:

$$ATH_{SPL}(f) = 3.64 \cdot \left(\frac{f}{1000}\right)^{-0.8} - 6.5 \cdot e^{-0.6\left(\frac{f}{1000}-33\right)^2} + 10^{-3} \cdot \left(\frac{f}{1000}\right)^4, [\text{дБ}] \quad (7.3)$$

где f – частота в герцах;

- частотное маскирование (simultaneous masking), проявляющееся при воздействии маскира в течение некоторого времени одновременно на разных частотах сигнала;
- маскировка во временной области (temporal masking): если громкий звук маскирует следующий за ним слабый звук, то явление называется маскировкой вперед (post-masking), и оно может продолжаться от 5 мс до 300 мс в зависимости от силы и длительности маскира; маскировка назад (pre-masking), когда громкий звук маскирует звук, воспроизводимый до него, длительность которой составляет примерно 20 мс.

Выбор метода частотно-временного анализа неразрывно взаимосвязан с точностью вычисления психоакустической модели и общей концепцией построения системы обработки речевых и звуковых сигналов. Пакет дискретного вейвлет преобразования (ПДВП) или другими словами быстрое вейвлет преобразование [5] является важным инструментом частотно-временной декомпозиции звуковых и речевых сигналов для различных приложений систем мультимедиа [1,3]. Одно из достоинств обработки сигналов на основе ПДВП, согласованного со шкалой критических частотных полос, является возможность редактирования, как статистической избыточности, так и перцептуальной избыточности в сигнале, что позволяет проектировать системы компрессии с высоким качеством реконструированного сигнала. Применение многополосного вейвлет преобразования на основе гиперкомплексной арифметики – кватернионов для целей построения встраиваемых мультимедиа систем с перцептуальной обработкой акустической информации позволяет минимизировать ошибки арифметики с фиксированной запятой.

Использование WDFT [6,7] в одноканальных перцептуальных системах редактирования шума в речи и компрессии звуковых сигналов одновременно как базис для метода спектрального взвешивания и психоакустической модели позволяет осуществлять обработку в критических частотных полосах более точно в контексте психоакустического моделирования по сравнению со схемами на базе ДПФ.

Субъективные оценки качества восстановленной речи в системах перцептуальной обработки на данных подходах выше. Это объясняется тем, что весь процесс обработки осуществляется в перцептуальном домене с неравномерной частотной шкалой и нет необходимости в преобразованиях между разными частотными шкалами, что приводит к упрощению архитектуры системы.

Цель данной главы – показать методологию построения психоакустической модели для трех способов субполосной обработки речевых и звуковых сигналов, когда ошибка аппроксимации критических частотных полос минимальна в области Барков, а также получить процедуру расчета маскирующих порогов в пространстве коэффициентов соответствующего преобразования.

7.2 ПДВП, согласованное со шкалой критических частотных полос

Итак, пусть $\{\varphi_n(t): z \in \mathbb{Z}\}$ определяет множество структур деревьев ПДВП и пусть $\{E \subset \{(l, n): 1 \leq l \leq L, 0 \leq n \leq 2^l\}\}$ представляет собой узлы дерева ПДВП, тогда отрезок $[0,1)$ разделяется на диадические интервалы: $I_{l,n} = [n2^{-l}, (n+1)2^{-l}]$, которые соответствуют специфическому множеству узлов E . В частности, $\{\varphi_{l,n,k}(t): (l, n) \in E, k \in \mathbb{Z}\}$, где $\varphi_{l,n,k}(t) \triangleq 2^{-\frac{l}{2}}\varphi_{l,n}(2^{-l}t - k)$ является базовой формой в пространстве сигнала $\overline{\text{span}}\{\varphi_{0,0}(t - k): k \in \mathbb{Z}\}$. Узел $(l, n) \in E$ дерева ПДВП ассоциируется с частотной полосой, у которой центральная частота и полоса пропускания приблизительно задаются следующими соотношениями:

$$f_{l,n} = 2^{-l}(GC^{-1}(n) + 0.5) \cdot \frac{f_s}{2}, \quad (7.4)$$

$$\Delta f_{l,n} = 2^{-l} \cdot \frac{f_s}{2}, \quad (7.5)$$

где $GC^{-1}(n)$ – обратный код перестановок Грея, f_s – частота дискретизации сигнала.

ПДВП реализуется на выбранной структуре дерева, поиск которой основывается на известном утверждении [8,9]: любая комбинация целых индексов $(l, n, k) \in \mathbb{Z}$, для которых вейвлеты сконцентрированы на двоичных интервалах $[n2^{-l}, (n+1)2^{-l}]$ из диапазона $[0, \infty)$, соответствует ортогональным базисам $\psi_{l,n,k}(t)$, $\varphi_{l,n,k}(t)$ из пространства $L^2(k)$. Утверждение доказывает существование множества структур ПДВП. Причем, ПДВП ассоциируется с алгоритмом выбора лучшей структуры преобразования $\{E \subset \{(l, n): 1 \leq l \leq L, 0 \leq n \leq 2^l\}\}$ из множества структур путем изменения и минимизации определенной меры качества.

Для того, чтобы получить аппроксимацию шкалы критических частотных полос с помощью ПДВП, необходимо осуществить декомпозицию дерева ПДВП таким образом, чтобы рас-

стояние между центрами одной субполосы и другой субполосы составляло 1 Барк. Следует отметить, что ширина критических частотных полос $CBW(f)$ монотонно увеличивающаяся функция частоты (7.2). Для формирования низкочастотных полос требуется интенсивная декомпозиция ПДВП в сравнении с характером изменения дерева ПДВП для аппроксимации высокочастотных полос.

Итак, дано дерево ПДВП $(l, n) \in E_m$ и его вейвлет коэффициенты $X_{l,n,k}$. Интегральная перцептуально взвешенная ошибка аппроксимации шкалы критических частотных полос деревом $(l, n) \in E_m$ ПДВП в области Барков может быть определена следующим образом:

$$Q_E = \frac{1}{L} \sum_{\substack{\text{для} \\ \forall (l,n) \in E_m}} [\widehat{CBW}_{z_w}(z) - \widehat{CBW}_{E_m}(z_{(l,n)})]^2 \cdot \widehat{W}(z). \quad (7.6)$$

Здесь ширина критических частотных полос $\widehat{CBW}(z)$ в Гц как функция центральных частот соседних критических частотных полос заданных в Барках, т.е.

$$\widehat{CBW}(z) = CBW(F^{-1}(z)), [\text{Гц}] \quad (7.7)$$

$\widehat{CBW}_{z_w}(z)$ определяет шкалу критических частотных полос в модели Zwicker [4], $\widehat{CBW}_{E_m}(z_{(l,n)})$ - аппроксимация критических частотных полос деревом ПДВП $(l, n) \in E_m$, центр $z_{(l,n)}$ в Барках полосы (l, n) дерева ПДВП E_m вычисляется для центральной частоты $f_{(l,n)}$, заданной в Гц, как $z_{(l,n) \in E_m} = F(f_{(l,n)})$, где F - преобразование (7.1). Перцептуальная взвешивающая функция $\widehat{W}(z)$, учитывающая определенные частотные свойства наружного и среднего уха. Она задает распределение ошибки аппроксимации шкалы критических частотных полос меньше в области средних частот по сравнению с низкочастотным и высокочастотным диапазонами, и определяется в шкале дБ, как функция частоты [10]:

$$W_{\text{дБ}}(f) = -0.6 \times 3.64(10^{-3}f)^{-0.8} + 6.5 \times \exp(-0.6 \times (10^{-3}f - 3.3)^2) - 10^{-3}(10^{-3}f)^4, \quad (7.8)$$

а также $\widehat{W}(z)$ может быть переопределена для Барков, как

$$\widehat{W}(z) = \widehat{W}(F(f)) = W(F^{-1}(z)) = W(f), \quad (7.9)$$

где $W(f) = 10^{(W_{\text{дБ}}(f)/20)}$.

Минимизация ошибки Q_E (7.6) может позволить автоматизировать процесс построения оптимального дерева ПДВП $(l, n) \in E_{CB}$ для шкалы критических частотных полос.

На рис. 7.1 показано дерево ПДВП (Critical Band Wavelet Packet Decomposition (CB - WPD)) [3,11], полученное эмпирически, которое осуществляет разделение частотного интервала сигнала на полосы, согласно критической шкале частот: CB - WPD: $(l, n) \in E_{CB}$, $l = \overline{0,8}$, где E_{CB} - обозначает множество узлов дерева ПДВП соответствующего CB - WPD. Дерево CB - WPD делит частотный диапазон [0 - 22,05 кГц] на 25 неравномерных полос $CBW(f)$, т.е. на 25 барков. Корневой узел $(l, n) = (0,0)$ данного дерева соответствует всему частотному диапазону сигнала. Каждый внутренний узел дерева $(l, n) \in E$, названный узлом предка, делится на два потомка: 1-й потомок и 2-й потомок, ассоциируемые соответственно с высокочастотной

и низкочастотной фильтрацией, выходные сигналы (вейвлет коэффициенты) которых децимируются в соотношении 2:1:

$$X_{l,n,k}(t) = \langle x(t), \varphi_{l,n,k}(t) \rangle, (l, n) \in E_{CB}, k \in \mathbb{Z}. \quad (7.10)$$

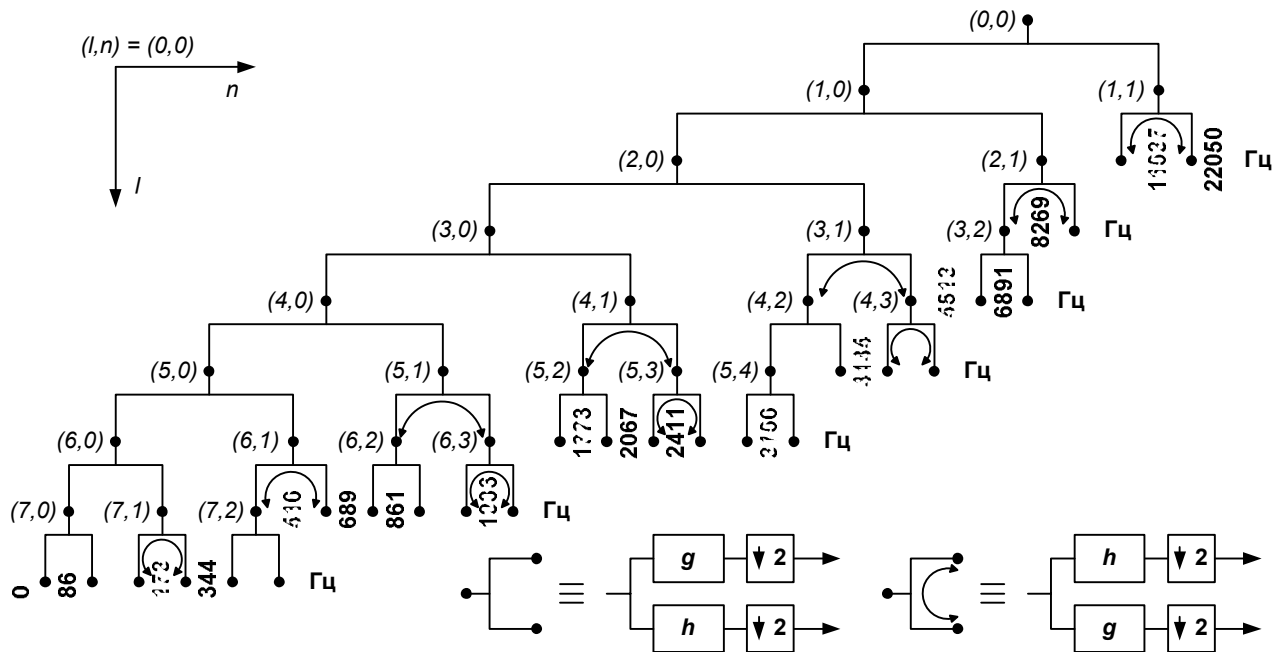


Рис. 7.1 – Структура критического дерева ПДВП $CB - WPD: (l, n) \in E_{CB}, l = \overline{0,8}$

Для систем перцептуальной обработки широкополосных речевых сигналов предлагается следующее критическое дерево ПДВП $CB-WPD$ [2,11]: $(l, n) \in E_{CB}, l = \overline{0,7}$ (см. рис. 7.2), где ширина полосы анализируемого сигнала равна 16 кГц и обработка вводится в 24 барках.

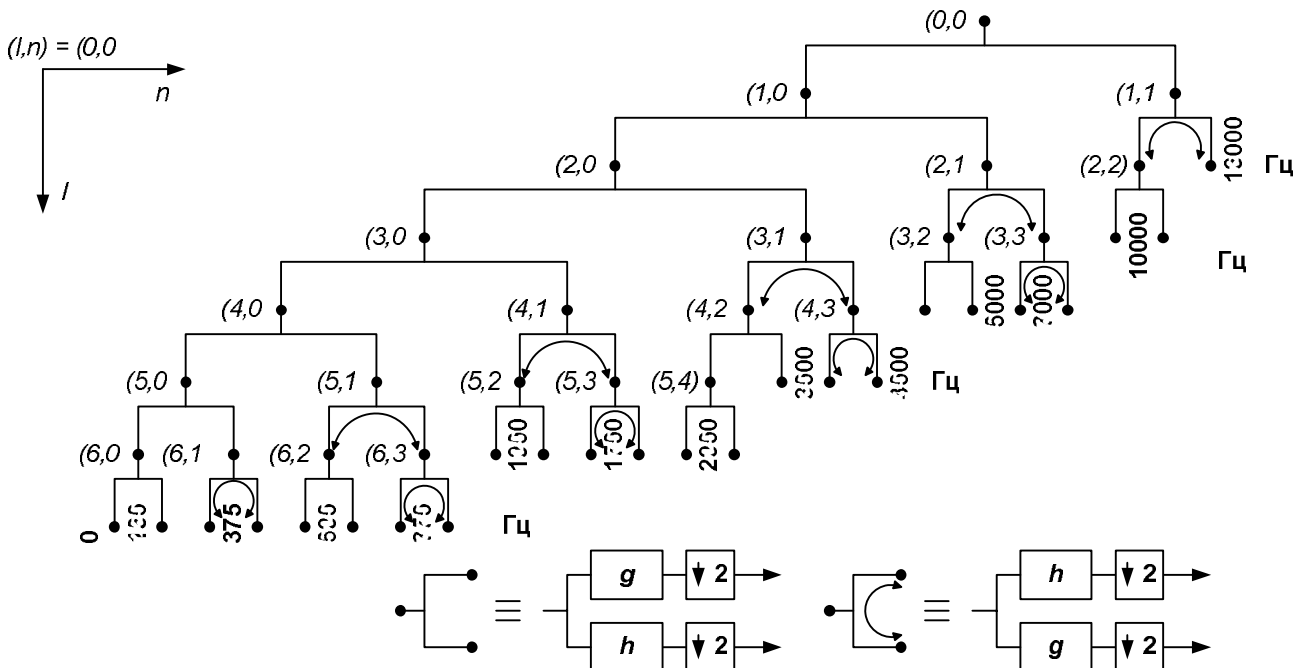


Рис. 7.2 – Структура критического дерева ПДВП $CB - WPD: (l, n) \in E_{CB}, l = \overline{0,7}$

Разрешающая способность человеческого уха ограничивает длину анализируемого фрейма в пределах 5-10 мс для области верхних частот и 100 мс для нижних частот. Выбор фильтра прототипа преобразования, длины его вейвлет функции, в общем случае зависят от максимального размера окна обработки (временного разрешения) и (нулевых) моментов [5]. Использование вейвлет функций семейства Добеши позволяет обеспечить хорошую частотную избирательность, которая увеличивается с числом масштабных уровней в дереве преобразования.

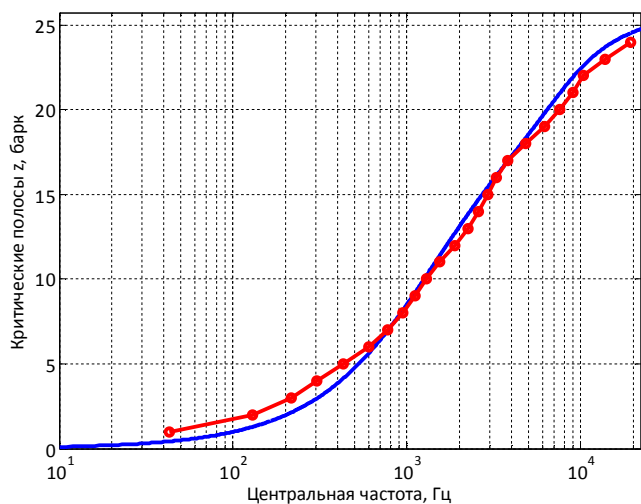


Рис. 7.3 – Аппроксимация центральных частот $CB - WPD: (l, n) \in E_{CB}, l = \overline{0,8}$

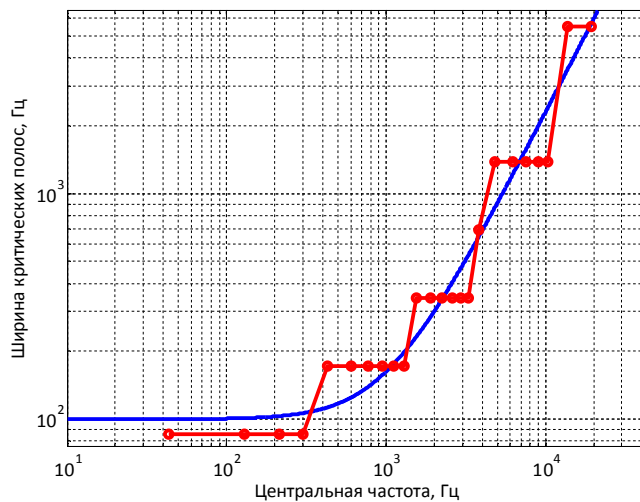


Рис. 7.4 – Аппроксимация ширины критических частотных полос $CB - WPD: (l, n) \in E_{CB}, l = \overline{0,8}$

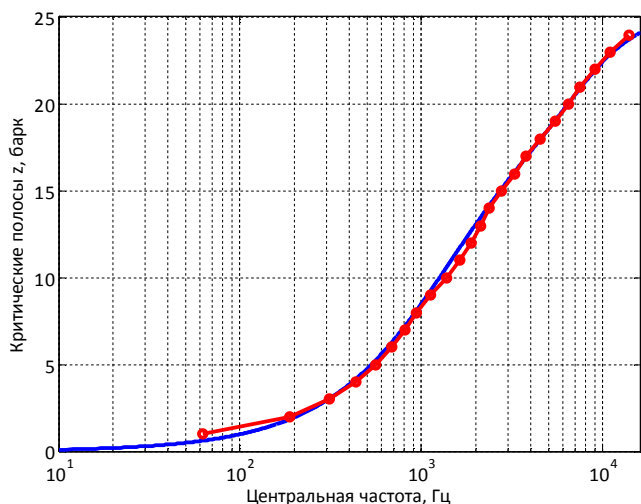


Рис. 7.5 – Аппроксимация центральных частот $CB - WPD: (l, n) \in E_{CB}, l = \overline{0,7}$

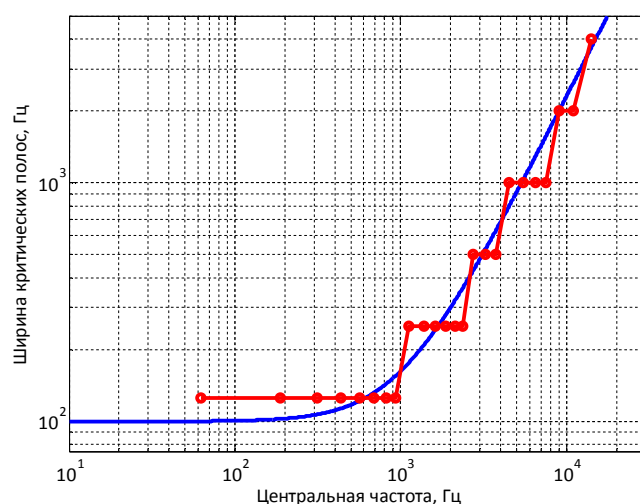


Рис. 7.6 – Аппроксимация ширины критических частотных полос $CB - WPD: (l, n) \in E_{CB}, l = \overline{0,7}$

На рис. 7.3 и рис. 7.4 показаны аппроксимации центральной частоты и ширины каждой частотной полосы критической шкалы частот соответственно деревом ПДВП, структура которого приведена на рис. 7.1. Аналогичные результаты для дерева ПДВП (см. рис. 7.2) иллюстрируются на рис. 7.5 и рис. 7.6. Здесь непрерывная линия соответствует критической шкале частот-

ных полос согласно модели Zwicker, а кружки – декомпозиции дерева ПДВП. Сопоставление данного результата с другими структурами дерева $CB - WPD$ [12,13] дает определенный выигрыш в величине ошибки Q_E (от 0 до 4 дБ), а также неформальные тесты прослушивания показывают лучшее восприятие восстановленных сигналов в перцептуальных системах компрессии и редактирования шумов.

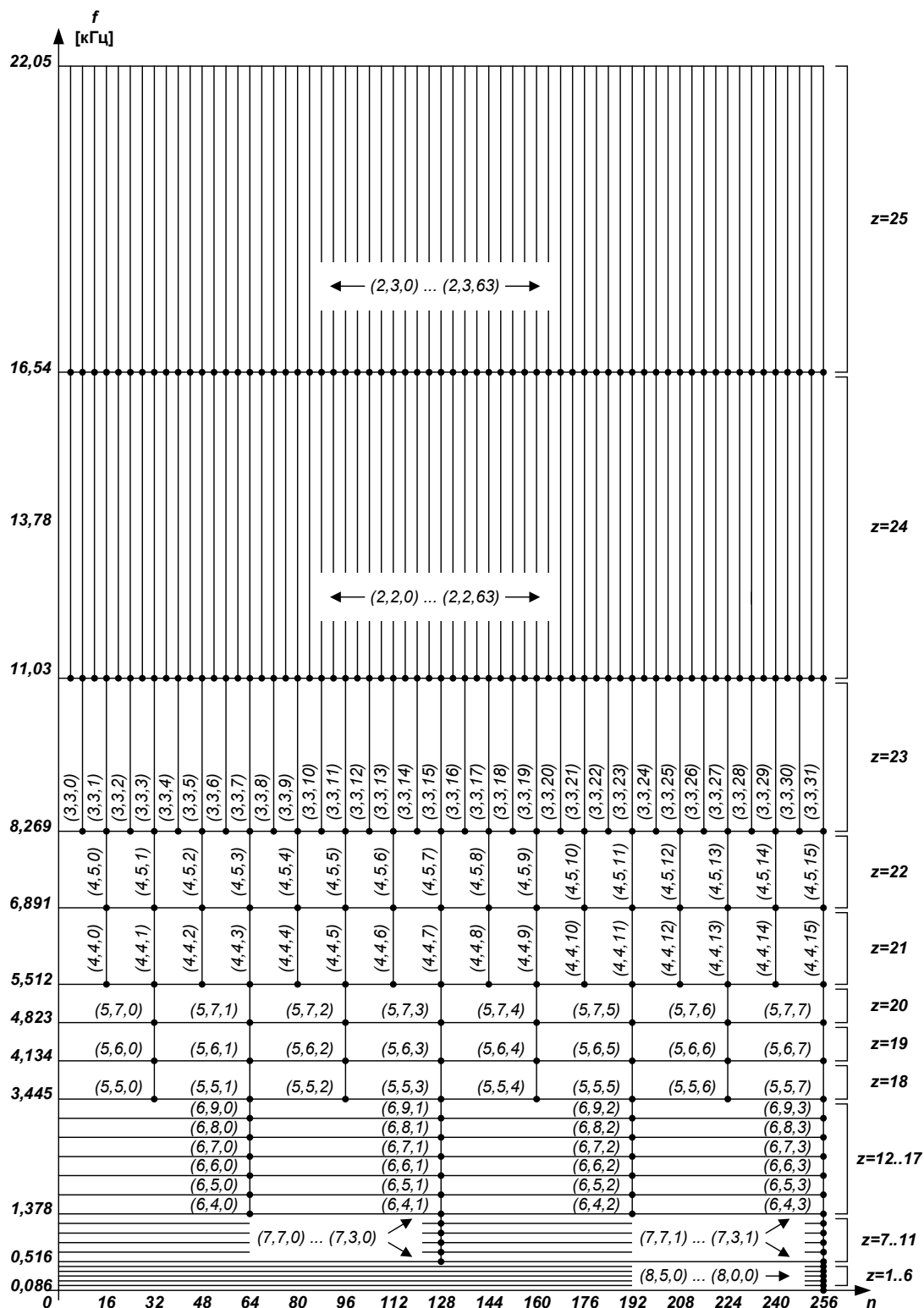


Рис. 7.7 – Частотно-временной план структуры дерева ПДВП
 $CB - WPD: (l, n) \in E_{CB}, l = \overline{0,8}, f_s = 44.1 \text{ кГц}$

Частотно-временной план для структуры дерева ПДВП $CB - WPD: (l, n) \in E_{CB}, l = \overline{0,8}$, $f_s = 44.1$ кГц (рис. 7.1) [3,11] представлен на рис. 7.7. Ширина каждой клеточки есть длина фрейма, и определяется как $F_l = 2^l$ ($F_{min} = 2$ отсчетов и $F_{max} = 256$ отсчетов). Следовательно, длина анализируемого окна равна $W = (P - 1)(F_{l-1}) + 1$ отсчетов. Для первого уровня $l = 1$ преобразования определяющей является область верхних частот и длина окна $W = 40$ отсчетов при длине фильтра прототипа $P = 40$. Для уровня $l = 8$ преобразования наибольшая частотная разрешающая способность в области нижних частот, а окно $W = 9946$ отсчетов.

В табл. 7.1 приведены показатели аппроксимации критических частотных полос структурами деревьев ПДВП $(l, n) \in E_{CB}, l = \overline{0,8}$ и $(l, n) \in E_{CB}, l = \overline{0,7}$ соответственно.

Табл. 7.1 – Аппроксимации критических частотных полос деревом ПДВП $(l, n) \in E_{CB}, l = \overline{0,8}$, $f_s = 44.1$ кГц и $(l, n) \in E_{CB}, l = \overline{0,7}$, $f_s = 32$ кГц

№ Барка	Узел		Кол. вейв-лет коэф.	Параметры полосы (Гц)		
	l	n		Нижняя	Центр	Верхняя
1	8	0	1	0.0	43.1	86.1
2	8	1	1	86.1	129.2	172.3
3	8	2	1	172.3	215.3	258.4
4	8	3	1	258.4	301.5	344.5
5	8	4	1	344.5	387.6	430.7
6	8	5	1	430.7	473.7	516.8
7	7	3	2	516.8	602.9	689.1
8	7	4	2	689.1	775.2	861.3
9	7	5	2	861.3	947.5	1033.6
10	7	6	2	1033.6	1119.7	1205.9
11	7	7	2	1205.9	1292.0	1378.1
12	6	4	4	1378.1	1550.4	1722.7
13	6	5	4	1722.7	1895.4	2068.2
14	6	6	4	2068.2	2240.0	2411.7
15	6	7	4	2411.7	2584.0	2756.3
16	6	8	4	2756.3	2928.5	3100.8
17	6	9	4	3100.8	3273.0	3445.3
18	5	5	8	3445.3	3789.8	4134.4
19	5	6	8	4134.4	4478.9	4823.4
20	5	7	8	4823.4	5168.0	5512.5
21	4	4	16	5512.5	6201.6	6890.6
22	4	5	16	6890.6	7579.7	8268.8
23	3	3	32	8268.8	9646.9	11025.0
24	2	2	64	11025.0	13781.2	16537.5
25	2	3	64	16537.5	19293.8	22050.0

№ Барка	Узел		Кол. вейв-лет коэф.	Параметры полосы (Гц)		
	l	n		Нижняя	Центр	Верхняя
1	7	0	1	0.0	62.5	125.0
2	7	1	1	125.0	187.5	250.0
3	7	2	1	250.0	312.5	375.0
4	7	3	1	375.0	437.5	500.0
5	7	4	1	500.0	562.5	625.0
6	7	5	1	625.0	687.5	750.0
7	7	6	1	750.0	812.5	875.0
8	7	7	1	875.0	937.5	1000.0
9	6	4	2	1000.0	1125.0	1250.0
10	6	5	2	1250.0	1375.0	1500.0
11	6	6	2	1500.0	1675.0	1750.0
12	6	7	2	1750.0	1875.0	2000.0
13	6	8	2	2000.0	2125.0	2250.0
14	6	9	2	2250.0	2375.0	2500.0
15	5	5	4	2500.0	2750.0	3000.0
16	5	6	4	3000.0	3250.0	3500.0
17	5	7	4	3500.0	3750.0	4000.0
18	4	4	8	4000.0	4500.0	5000.0
19	4	5	8	5000.0	5500.0	6000.0
20	4	6	8	6000.0	6500.0	7000.0
21	4	7	8	7000.0	7500.0	8000.0
22	3	4	16	8000.0	9000.0	10000.0
23	3	5	16	10000.0	11000.0	12000.0
24	2	3	32	12000.0	14000.0	16000.0

7.3 Процедура расчета порогов маскирования в вейвлет области

Пусть дано $CB - WPD: (l, n) \in E_{CB}$, частотно-временной план дерева ПДВП E_{CB} , и коэффициенты $X_{l,n,k}$. Процедура расчета порогов маскирования следующая [14]:

- вычислить спектральную энергию барка:

$$A_{CB}(z) = \sum_{k=0}^{K-1} X_{z,k}^2, \quad (7.11)$$

- оценить тональность сигнала в каждой критической частотной полосе и значения индексов $a_{tmn}(z)$ и $a_{nmn}(z)$ уменьшения спектральной энергии барка соответственно для тоновых и шумовых маскеров:

- индекс $a_{tmn}(z)$, который оценивает отношение маскирования тоном шума, задается

так:

$$a_{tmn}(z) = -0.275 \cdot z - 15.025 \text{ [дБ]}, \quad (7.12)$$

- индекс маскирования шумом шума $a_{nmn}(z)$ оценивается как константа

$$a_{nmn}(z) = -25 \text{ [дБ]}, \quad (7.13)$$

так как ПДВП уже внесло некоторое спектральное перекрытие;

- среднее значение тональности маскеров в каждой критической полосе определяется маскирующим индексом:

$$a_{CB}(z) = \eta \cdot a_{tmn}(z) + (1 - \eta) \cdot a_{nmn}(z) \text{ [дБ]}, z = \overline{1,25}, \quad (7.14)$$

где η - тональный коэффициент:

$$\eta = \min(SFM(z)_{\text{дБ}} / SFM_{\text{дБmax}}, 1), \quad (7.15)$$

где $SFM_{\text{дБ}}$ – мера спектральной пологости [15], рассчитываемая как

$$SFM(z) = \frac{(\prod_{k=0}^{K-1} X_{z,k}^2)^{\frac{1}{K}}}{\frac{1}{K} \sum_{k=0}^{K-1} X_{z,k}^2}, \quad (7.16)$$

$SFM_{\text{дБmax}}$ – максимальное значение меры пологости спектра. Для заданного фильтра прототипа $SFM_{\text{дБmax}} = -25$ дБ;

- найти спектральную энергию Барка z с учетом тональности сигнала:

$$D_{CB}(z) = 10 \cdot \log \left(A_{CB}(z) \cdot 10^{\frac{a_{CB}(z)}{10}} \right) \text{ [дБ]}. \quad (7.17)$$

- вычислить разброс энергии барка $C_{CB}(z)$ как свертку $D_{CB}(z)$ с функцией разброса $B(z)$ в каждой критической частотной полосе z :

$$C_{CB}(z) = 10 \cdot \log \left(\frac{1}{K} \sum_{k=1}^{25} 10^{\frac{D_{CB}(k)}{10}} \cdot 10^{\frac{B(z-k)}{10}} \right) \text{ [дБ]}, z = \overline{1,25}, \quad (7.18)$$

где функция разброса $B(z)$ вычисляется как

$$B(z) = a + \frac{v + u}{2} \cdot (z + c) - \frac{v - u}{2} \cdot \sqrt{d + (z + c)^2}, \quad (7.19)$$

а параметры функции a, v, u, d, c приведены в первой строке табл. 7.2.

Табл. 7.2 – Параметры функции разброса

Функция разброса	v	u	d	c	a
Барк шкала	30 дБ/барк	-25 дБ/барк	0.3	0.05	15
Временная шкала	$0.0825 \text{ дБ}/F_{min}^*$	$-0.0412 \text{ дБ}/F_{min}^*$	0.3	0.157	$0.032/F_{min}^*$

F_{min}^* – минимальная длина анализируемого фрейма .

- вычислить пороги маскирования во временной области:

аналогично, как и в частотном маскировании, во временном маскировании уже присутствуют некоторые элементы перекрытия в АЧХ, обусловленные деревом ПДВП. Предполагается, что временное маскирование аддитивно к сигналу и определяется через коэффициенты ПДВП в каждой критической частотной полосе z (см. рис. 7.8). Максимальное временное разрешение для ПДВП имеет место в критических частотных полосах верхних частот, которые имеют минимальную протяженность по времени $F_{min}^* = 2$ отсчета или 0.0454 мс. Временная функция разброса $B(k)$ в вейвлет области задается как

$$B(k) = a + \frac{v + u}{2} \cdot (k + c) - \frac{v - u}{2} \cdot \sqrt{d + (k + c)^2} \text{ [дБ]}, \quad (7.20)$$

параметры которой вдоль оси времени определяются как $v = 20 \text{ дБ}/\text{мс} = 0.0825 \text{ дБ}/F_{min}^*$ и $u = 20 \text{ дБ}/\text{мс} = -0.0412 \text{ дБ}/F_{min}^*$ (см. табл. 7.2 строка 2). На рис. 7.8 схематически показано проявление временного маскирования.

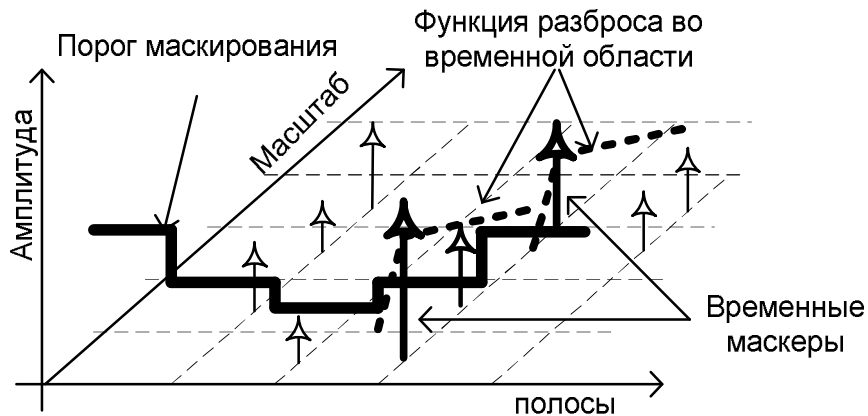


Рис. 7.8 – Схема временного маскирования в соответствующей критической частотной полосе z

- вычислить энергию вейвлет коэффициентов в каждой критической частотной полосе z :

$$E_z(k) = X_{z,k}^2, \quad k = \overline{1, K-1}, \quad z = \overline{1, 25}; \quad (7.21)$$

- определить временную функцию разброса энергии в каждой критической частотной полосе z , как свертку $E_z(k)$ и функции разброса $B(k)$:

$$F_z(m) = \frac{1}{K} \sum_{k=0}^{K-1} E_z(k) \cdot 10^{\frac{B(K-k)}{10}}, m = \overline{1, K-1}; \quad (7.22)$$

- найти временной фактор маскирования в полосе z , как результат сравнения величин:

$$F_z(m) \geq E_z(m), k = \overline{1, K-1}. \quad (7.23)$$

Если данное соотношение выполняется, то в соответствующей критической частотной полосе имеет место временное маскирование, в противном случае его нет.

- оценить частотно-временной маскирующий порог $M_{CB}(z)$ в каждой критической частотной полосе:

$$M_{CB}(z) = C_{CB}(z) \cdot \max\left(\frac{F_z(k)}{E_z(k)}, 1\right) \text{ [дБ]}, k = \overline{1, K-1}. \quad (7.24)$$

- вычислить значение маскирующего порога $T_{CB}(z)$ в соответствующей критической частотной полосе из сравнения временно-частотного маскирующего порога $M_{CB}(z)$ с минимальным значением абсолютного порога слышимости $ATH(z)$ (7.3):

$$T_{CB}(z) = \max(ATH(z), M_{CB}(z)) \text{ [дБ]}. \quad (7.25)$$

Предложенная процедура расчета параметров психоакустической модели в области вейвлет коэффициентов нашла применение при реализации систем кодирования звуковых и широкополосных речевых сигналов [2], а также используется в моделировании переходных компонент сигнала на основе согласованной подгонки с адаптивным словарем, формируемым структурой дерева ПДВП в параметрических кодерах [16]. Ведение обработки сигнала и реализация психоакустической модели в одной области исключает необходимость перерасчета параметров модели, что позволяет исключить дополнительную погрешность.

Программа вычисления функции разброса Шрейдера.

```
function [sf] = spreadf(z,mode)
%
% Implemented Model for Spreading function
%
if strcmp(mode,'schr'), % Schroeder et al.
a = [15.81];
c = [ 0.47];
d = [ 1];
u = [ -10]; % dB/Bark
v = [ 25]; % dB/Bark
elseif strcmp(mode,'bark'), % Bark scale
a = [27.39];
c = [ 0.05];
d = [ 0.3];
u = [ -25]; % dB/Bark
v = [ 30]; % dB/Bark
elseif strcmp(mode,'temp'), % Temporal scale
a = [ 7.75];
c = [ 0.19];
```

```

d = [ 0.3];
u = [ -10]; % dB/Fmin
v = [ 20]; % dB/Fmin
end;
sf = a + (v + u) / 2 * (z + c) - (v - u) / 2 * sqrt(d + (z + c) .^ 2);

```

Программа вычисления психоакустической модели на базе вейвлетного пакета.

```

[filename, pathname] = uigetfile('*.wav','Open File ...');
[x,Fs,bits] = wavread(filename);

start = 0;
tree.N = [256:259 130:135 68:73 37 19:23 6:7];
tree.wname = 'db6';

while start < length(x),
    new_x = x(start:start+2^fix(log2(max(tree.N))-1));
    [data,tree] = wpdec(new_x,tree);
    tree_all.N = [2^fix(log2(max(tree.N))):2^ceil(log2(max(tree.N))-1)];
    tree_all.wname = tree.wname;
    [data_all,tree_all] = wpdec(new_x,tree_all);
%-----
% Critical Band Analysis
%-----
    beg_all = 1; PSD_all = [];
    for i = 1:tree_all.length,
        PSD_all = [PSD_all ...
            (wkeep(data_all(beg_all:beg_all+tree_all.sizes(i)-1), 1)) .^2];
        beg_all = beg_all + tree_all.sizes(i);
    end;

    beg = 1; PSD = []; t1 = [];
    for z = 1:length(tree.N),
        ind(z) = 2^(fix(log2(max(tree.N))-fix(log2(tree.N(z)))));
%-----
% Power Spectrum
%-----
        PSD = [PSD ...
            10*log10((wkeep(data(beg:beg+tree.sizes(z)-1), ind(z))) .^2) ];
        beg = beg + tree.sizes(z);
    end;

    PSD = PSD + 90;

    beg = 1; Af = []; Df = [];
    for z = 1:length(tree.N),
%-----
% Bark Energy Spectrum
%-----
        A(z) = 10*log10(sum(10.^( PSD(beg:beg+ind(z)-1)/10)));
        for i = 1:ind(z), Af = [Af A(z)]; end;
%-----

```

```

% Spectrum Flatness Measure
%-----
SFM(z) = ( (prod(PSD_all(beg:beg+ind(z)-1))^(1/ind(z))) / ...
(sum(PSD_all(beg:beg+ind(z)-1)/ind(z))
);
if SFM(z) ~= 0,
    SFM(z) = 10*log10(SFM(z));
else
    SFM(z) = -inf;
end;
beg = beg + ind(z);
%-----
% Paramets tau is tonality coefficient
%-----
tau(z) = min(SFM(z)/ -60 , 1);
a.tmn(z) = -0.275*(z-1) - 15.025; %dB
a.nmn(z) = -9.0; %dB
a.all(z) = tau(z)*a.tmn(z) + (1-tau(z))*a.nmn(z);
%-----
% Schifted Bark Energy Spectrum
%-----
D(z) = 10*log10(10^(A(z)/10)*10^(a.all(z)/10));
for i = 1:ind(z), Df = [Df D(z)]; end;
end;

Cf = [];
for z = 1:length(tree.N),
%-----
% Linear version of spreading function
%-----
Bz(z,:) = 10.^(spreadf(1-z:length(tree.N)-z,'bark')/10);
%-----
% Spreading of Bark Energy
%-----
C(z) = 10*log10(1/(ind(z)).*sum((10.^(D/10)).*Bz(z,:)));
for i = 1:ind(z), Cf = [Cf C(z)]; end;
end;
%-----
% Temporal energy sequence
%-----
Ez = PSD; beg = 1;
for z = 1:length(tree.N),
    for n = 1:ind(z),
        Bn = 10.^(spreadf(max(ind)/ind(z)- ...
            max(ind)/ind(z)*n:max(ind)/ind(z):max(ind)- ...
            max(ind)/ind(z)*n,'temp')/10);
        Fz(beg+n-1) = 10*log10(sum(10.^(Ez(beg:beg+ind(z)-1)/10).*Bn));
    end;
    beg = beg + ind(z);
end;

beta = []; beg = 1;
for z = 1:length(tree.N),
    temp = (Fz(beg:beg+ind(z)-1))./abs(Ez(beg:beg+ind(z)-1));

```

```

        temp = 1/(ind(z))*sum(temp);
        if temp < 1, temp = 1; end;
        beta = [beta temp];
        beg = beg + ind(z);
    end;

    Mf = []; beg = 1;
    for z = 1:length(tree.N),
        M(z) = 10*log10(10.^(C(z)/10)*10.^(beta(z)/10));
        for i = 1:ind(z), Mf = [Mf M(z)]; end;
        beg = beg + ind(z);
    end;

    f = round(Fs/(2*tree.length)):round(Fs/(2*tree.length)): ...
        round(Fs/(2*tree.length))*(tree.length);

    for i = 1:length(f),
        M_b(i) = 13 * atan(0.00076*f(i))+3.5*atan((f(i)/7500).^2);
    end;
    T = threshold('b');
    for i = 1:tree.length,
        Tg(i) = 10*log10( 10.^(T(i,1)/10) + 10.^(Mf(i)/10) );
    end;
    Ml = 10*log10(M);

    figure(1);
    subplot(211);
    plot(new_x);
    axis([1 length(new_x) min(x) max(x)]);
    grid on;
    subplot(212);
    plot(T(:,2),T(:,1),'k-');
    hold on;
    plot(M_b,Ez,'b:');
    plot(M_b,Mf,'r')
    grid on
    axis([1 25 -30 100])
    hold off
    start = start + 256;
end;

```

7.4 Перцептуальная обработка в домене 4-х полосного параунитарного банка фильтров на основе кватернионов

7.4.1 Многополосные вейвлеты

Многополосные вейвлеты представляют собой обобщение диадического вейвлет преобразования, и которые также могут быть представлены в виде структуры дерева. Такой многополосный банк фильтров называется M – полосное вейвлет преобразование. В данном случае используется M фильтров с импульсными характеристиками $h_i(l)$, $l = 0 \dots L - 1$, в то время как в диадическом вейвлет преобразовании применяется только два фильтра. Масштабная

функция получается в результате решения следующего разностного уравнения для двух масштабов с использованием только фильтра нижних частот

$$\psi_0 = \sqrt{M} \sum_{l=0}^{L-1} h_0(l) \cdot \psi_0(Mt - l). \quad (7.26)$$

Далее на основании масштабной функции могут быть определены вейвлетные функции

$$\psi_i = \sqrt{M} \sum_{l=0}^{L-1} h_i(l) \cdot \psi_0(Mt - l), \quad i = 1 \dots M - 1. \quad (7.27)$$

Сдвигая во времени и масштабируя данные вейвлеты можно с помощью их представить произвольные интегрируемые квадратичные функции. Следует отметить, что M – полосное вейвлет преобразование имеет ряд преимуществ перед диадическим аналогом. Во-первых, эффективность кодирования здесь выше. Во-вторых, 2-х канальный ортогональный банк фильтров не имеет линейной фазы, в то время как данное ограничение отсутствует для $M > 2$. Однако, проектирование и применение M – канального банка фильтров имеет определенные сложности, в то время как теория диадического вейвлет преобразования и соответствующих банков фильтров хорошо разработана. Следует отметить такой важный результат для многополосных вейвлетов на основе кватернионов как нечувствительность коэффициентов к квантованию касательно лестничных и решетчатых структур - структур без потерь, осуществляющих точную реконструкцию входного сигнала по схеме анализ-синтез [17-19]. Точное соотношение между энергиями во всем частотном диапазоне и в субполосах сильно упрощает теоретические выкладки и делает параунитарные банки фильтров (ПУБФ) полезными в применении к компрессии (кодированию) сигналов. Свойство параунитарности означает, что базовые функции соответствующих субполос банка фильтров являются ортогональными. Таким образом, данные структуры могут быть выбраны для реализации на арифметике с фиксированной запятой.

7.4.2 Решетчатая структура M -канального параунитарного банка фильтров с линейной фазой на кватернионах

Более подходящим описанием структуры ПУБФ анализа является $M \times M$ полифазная матрица $\mathbf{E}(z)$, которая будет параунитарной, если $\mathbf{E}^H(z^{-1})\mathbf{E}(z) = c\mathbf{I}_M$, где c – ненулевая константа и M обозначает число каналов в банке фильтров. Структуру ПУБФ получают путем конструирования его полифазной матрицы $\mathbf{E}(z)$ из некоторых элементарных математических преобразований, как правило, это планарные обороты Гивенса. Формально параунитарный банк фильтров анализа общего вида может быть описан следующим выражением:

$$\mathbf{E}(z) = \mathbf{R}_N \mathbf{\Lambda}(z) \mathbf{R}_{N-1} \mathbf{\Lambda}(z) \dots \mathbf{R}_1 \mathbf{\Lambda}(z) \mathbf{E}_0, \quad (7.28)$$

где \mathbf{E}_0 – общая ортогональная матрица; \mathbf{R}_i – ортогональные матрицы, определяющие передаточные характеристики банка фильтров ($i = 1 \dots N$); N – количество матриц \mathbf{R}_i ; $\mathbf{\Lambda}(z)$ – матрица задержки: $\mathbf{\Lambda}(z) = \text{diag}(z^{-1}, \mathbf{I}_{M-1})$, см. рис. 7.9.

Серьезной практической проблемой, связанной с реализацией ПУБФ на арифметике с фиксированной запятой, является потеря свойства перфективной реконструкции сигнала бан-

ком фильтров. Это обусловлено чувствительностью структуры ПУБФ к квантованию коэффициентов. В работе [20] был предложен новый подход к синтезу и конструированию ПУБФ на основе решетчатых структур, для которых базовым элементарным преобразованием является умножение кватернионов, т.е. умножитель гиперкомплексных чисел – кватернионов – есть основной «строительный блок» структур ПУБФ. Кватернион – это гиперкомплексное число размерности 4, которое описывается следующим выражением: $Q = q_1 + q_2i + q_3j + q_4k$, где i, j, k – мнимые единицы. Сложение кватернионов осуществляется аналогично сложению комплексных чисел, а умножение – по правилу перемножения многочленов. При этом умножение кватернионов не подчиняется переместительному закону умножения, но подчиняется сочетательному. Операции сложения и умножения кватернионов могут быть реализованы с использованием векторно-матричной арифметики. Для этого кватернионы представляются в виде 4-х мерных векторов. Умножение кватернионов в векторной форме производится по правилу умножения вектора на матрицу:

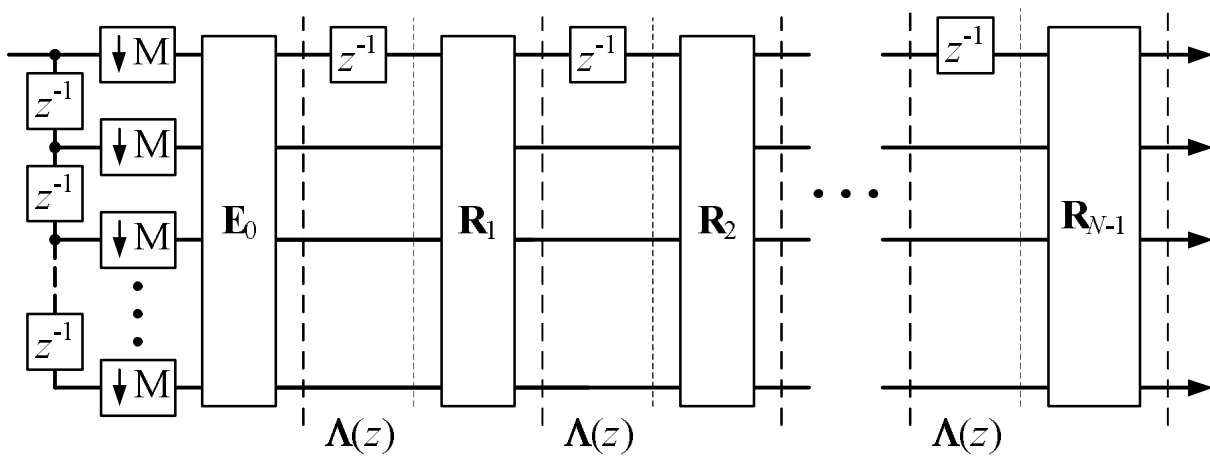


Рис. 7.9 – Решетчатая структура банка фильтров анализа

$$\begin{aligned}
 R = [r_1 \ r_2 \ r_3 \ r_4]^T = PQ &= \begin{bmatrix} p_1 & -p_2 & -p_3 & -p_4 \\ p_2 & p_1 & -p_4 & p_3 \\ p_3 & p_4 & p_1 & -p_2 \\ p_4 & -p_3 & p_2 & p_1 \end{bmatrix} \times \begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \end{bmatrix} \\
 &= \begin{bmatrix} q_1 & -q_2 & -q_3 & -q_4 \\ q_2 & q_1 & q_4 & -q_3 \\ q_3 & -q_4 & q_1 & q_2 \\ q_4 & q_3 & -q_2 & q_1 \end{bmatrix} \times \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix},
 \end{aligned} \tag{7.29}$$

где $\mathbf{M}^+(P)$ – левый оператор умножения для кватерниона P ; $\mathbf{M}^-(Q)$ – правый оператор умножения для кватерниона Q .

В случае использования умножителей кватернионов матрицы \mathbf{E}_0 и \mathbf{R}_i можно выразить через коэффициенты в виде кватернионов следующим образом:

$$\mathbf{E}_0 = \mathbf{M}^+(P_0)\mathbf{M}^-(Q_0), \mathbf{R}_i = \mathbf{M}^+(P_i), \tag{7.30}$$

где Q_0, P_i – соответствующие значения кватернионов.

Главное преимущество метода [19] не чувствительность 4-х и 8-ми канальных структур ПУБФ к квантованию коэффициентов матрицы умножения кватернионов. Следовательно, ПУБФ

на основе умножителей кватернионов является преобразованием без потерь. Банк фильтров синтеза (обратное преобразование) может быть получен путем инвертирования преобразования (7.28).

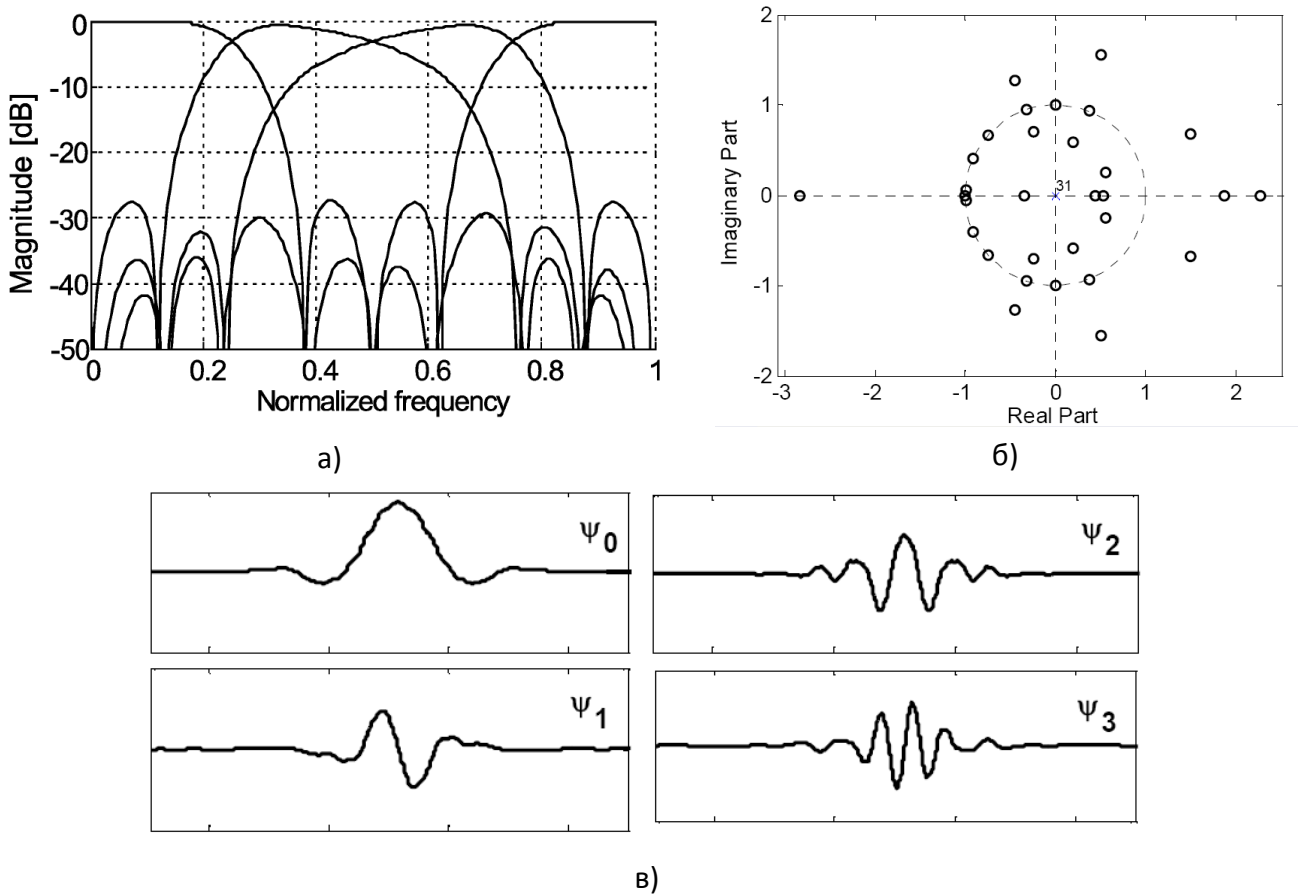


Рис. 7.10 – Характеристики 4-х полосного ПУБФ на кватернионах: а) амплитудно-частотная характеристика; б) карта нулей низкочастотного фильтра прототипа; в) масштабная функция и вейвлеты

Табл. 7.3 – Коэффициенты решетчатой структуры ПУБФ на кватернионах

Коэффициенты	Точное значение коэффициента	Квантованное значение коэффициента	Представление коэффициента в системе CSD	Коэффициенты	Точное значение коэффициента	Квантованное значение коэффициента	Представление коэффициента в системе CSD
$Q_{0,1}$	-0.7062053	-0.7500000	-2-0 + 2-2	$P_{3,2}$	-0.9990265	-1.0000000	-2-0
$Q_{0,2}$	0.7080071	0.7500000	+2-0 - 2-2	$P_{4,1}$	-0.9388232	-0.9375000	-2-0 + 2-4
$P_{0,1}$	0.1604117	0.1562500	+2-3 + 2-5	$P_{4,2}$	-0.3443996	-0.3750000	-2-1 + 2-3
$P_{0,2}$	0.9870502	0.9843750	+2-0 - 2-6	$P_{5,1}$	0.8957751	0.8750000	+2-0 - 2-3
$P_{1,1}$	-0.7985673	-0.7500000	-2-0 + 2-2	$P_{5,2}$	0.4445076	0.4375000	+2-1 - 2-4
$P_{1,2}$	-0.6019055	-0.6250000	-2-1 - 2-3	$P_{6,1}$	-0.1761618	-0.1875000	-2-2 + 2-4
$P_{2,1}$	-0.8506589	-0.8750000	-2-0 + 2-3	$P_{6,2}$	0.9843612	0.9843750	+2-0 - 2-6
$P_{2,2}$	0.5257179	0.5312500	+2-1 + 2-5	$P_{7,1}$	0.8211270	0.8750000	+2-0 - 2-3
$P_{3,1}$	0.0441147	0.0468750	+2-4 - 2-6	$P_{7,2}$	0.5707455	0.5625000	+2-1 + 2-4

Для аппроксимация критических частотных полос деревом пакета многополосных вейвлет используется 4-х полосный банк фильтров на квантерионах, амплитудно-частотная характеристика которого показана на рис. 7.10.

Значения точных и квантованных коэффициентов данного банка приведены в табл. 7.3 (нулевые коэффициенты в таблицу не включены). Проектирование четырех полосного банка фильтров на квантерионах велось так, чтобы получить максимальное ослабление в полосе подавления субполос. Факторизация (7.28) имеет 7 ступеней, таким образом импульсная характеристика фильтра прототипа имеет 32 отсчета. При этом, ослабление в полосе подавления сигнала субполос составляет -27 дБ, а эффективность кодирования для AR(1) модели сигнала равна 8,44 дБ.

7.4.3 Психоакустическая модель на основе 4-х полосного вейвлет преобразования на кватернионах

На рис. 7.11 показана структура критического дерева на основе 4-х полосного вейвлет преобразования на кватернионах $CB - WPD: (l, n) \in E_{CB}, l = \overline{0,4}, f_s = 44.1$ кГц, а в табл. 7.4 приведены численные результаты аппроксимации критических частотных полос.

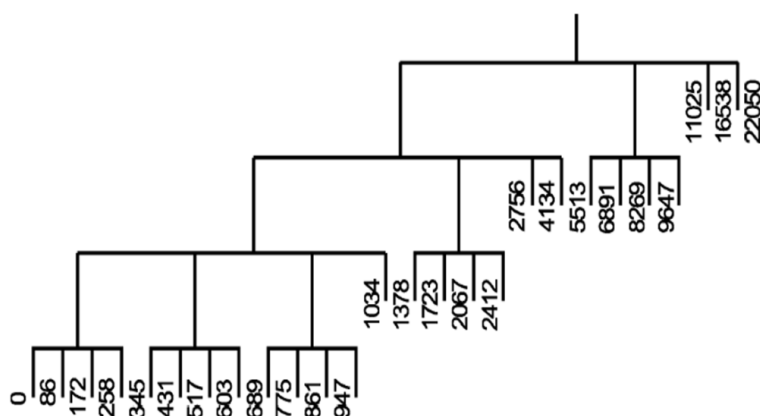


Рис. 7.11 – Структура критического дерева ПДВП $CB - WPD: (l, n) \in E_{CB}, l = \overline{0,4}, f_s = 44.1$ кГц

Табл. 7.4 – Аппроксимации критических частотных полос деревом ПДВП $(l, n) \in E_{CB}, l = \overline{0,4}$

Полосы	Параметры полосы [Гц]		
	Нижняя	Центральная	Верхняя
1	0	43	86
2	86	129	172
3	172	215	258
4	258	301	345
5	345	388	431
6	431	474	517
7	517	560	603
8	603	646	689
9	689	732	775
10	775	818	861
11	861	904	947
12	947	991	1034
13	1034	1206	1378

Полосы	Параметры полосы [Гц]		
	Нижняя	Центральная	Верхняя
14	1378	1550	1723
15	1723	1895	2067
16	2067	2239	2412
17	2412	2584	2756
18	2756	3445	4134
19	4134	4823	5513
20	5513	6202	6891
21	6891	7580	8269
22	8269	8958	9647
23	9647	10336	11025
24	11025	13781	16538
25	16538	19294	22050

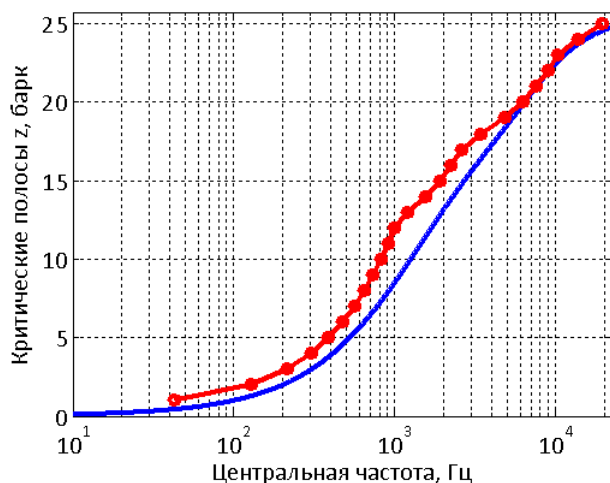


Рис. 7.12 – Аппроксимация центральных частот

$$CB - WPD: (l, n) \in E_{CB}, l = \overline{0,4}$$

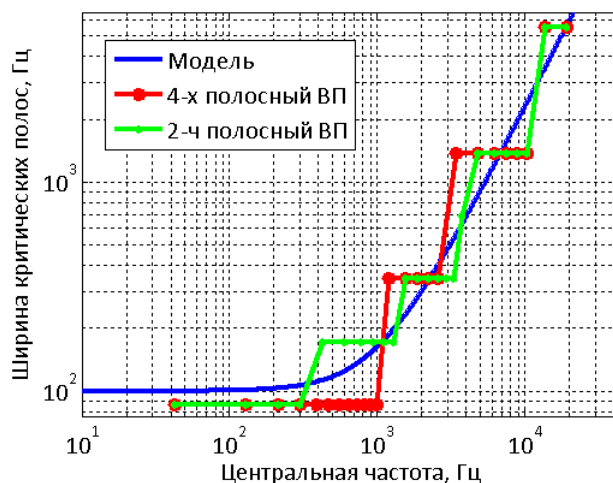


Рис. 7.13 – Аппроксимация ширины критических частотных полос

$$CB - WPD: (l, n) \in E_{CB}, l = \overline{0,4}$$

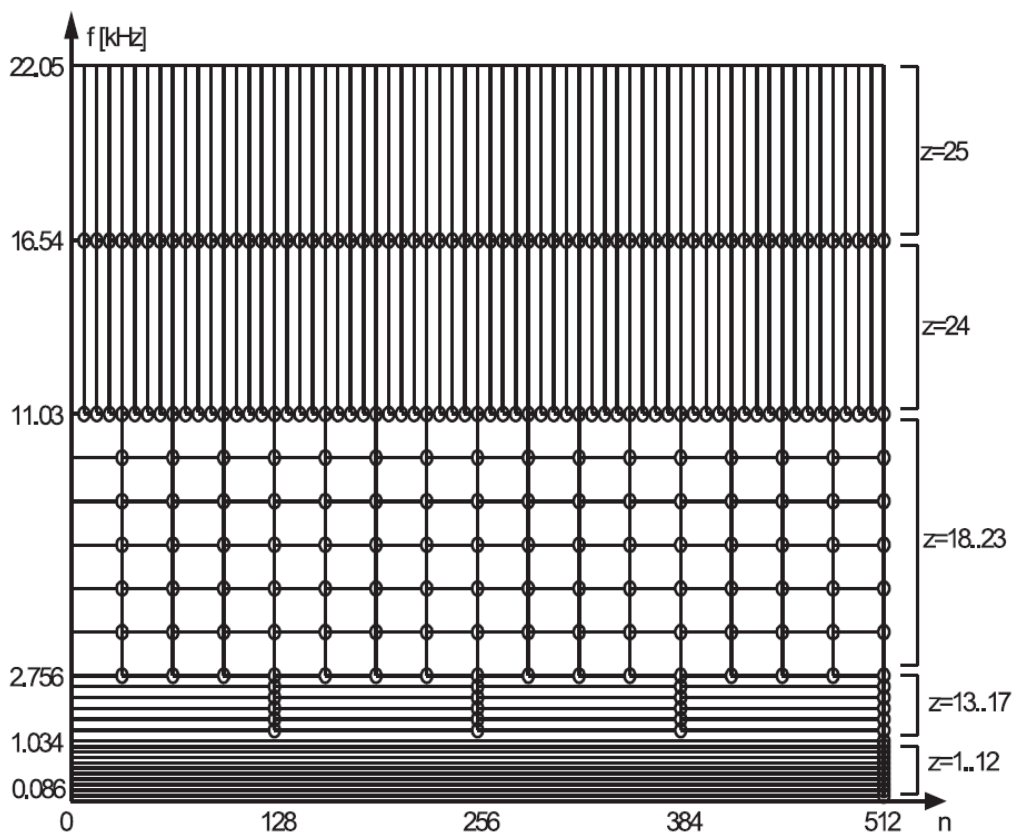


Рис. 7.14 – Частотно-временной план структуры дерева ПДВП

$$CB - WPD: (l, n) \in E_{CB}, l = \overline{0,8}, f_s = 44.1 \text{ кГц}$$

Из рис. 7.12 и рис. 7.13 хорошо видно, что аппроксимация центральных частот и ширины критических частотных полос деревом $CB - WPD$ на ветвях кватернионах: $(l, n) \in E_{CB}, l = \overline{0,4}$ удовлетворительная. Здесь непрерывная линия соответствует критической шкале частотных полос согласно модели Zwicker, а кружки – декомпозиции дерева ПДВП. Сопоставление данно-

го результата с другими структурами дерева $CB - WPD$ [12] дает определенный выигрыш в величине ошибки Q_E . Частотно-временной план структуры дерева ПДВП на вейвлет кватернионах изображен на рис. 7.14.

Процедура расчета порогов маскирования в области вейвлет кватернионов соответствует этапам вычисления аналогичной процедуры в домене коэффициентов двух-полосных вейвлетов.

7.5 Психоакустическая модель на основе WDFТ

7.5.1 Основные принципы WDFТ

WDFТ позволяет получить неравномерное z -преобразование конечной последовательности, используя фазовый фильтр $A(z)$. WDFТ последовательности $x(n)$ из N точек определяется по следующей формуле:

$$\hat{X}(z_k) = X(\hat{z}_k) = \sum_{n=0}^{N-1} x(n)\hat{z}_k^{-n}, k = 0, \dots, N - 1, \quad (7.31)$$

где \hat{z}_k – изображения равноотстоящих точек на единичной окружности в z -плоскости, получаемые из преобразования

$$\hat{z}_k^{-1} = e^{-j\frac{2\pi}{N}} \rightarrow \hat{z}_k^{-1} = A(z_k), k = 0, \dots, N - 1, \quad (7.32)$$

с фазовым звеном $A(z)$ произвольного порядка.

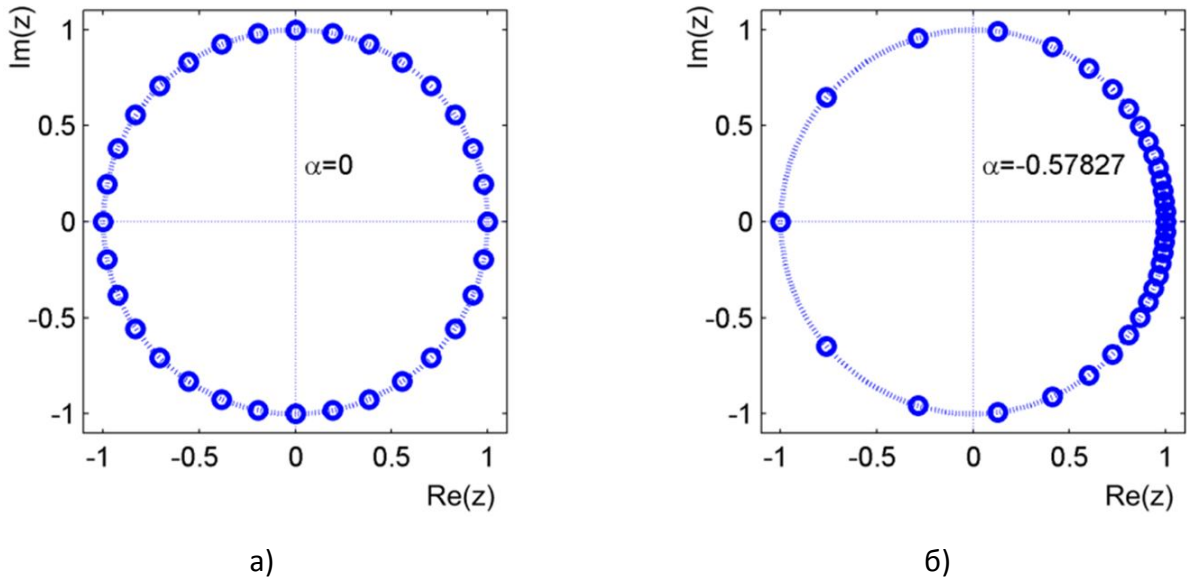


Рис. 7.15 – Расположение коэффициентов преобразования на единичной окружности в z -плоскости: ДПФ (а) и WDFТ (б)

Простейший вариант WDFТ основан на фазовом звене первого порядка с действительным коэффициентом [21]. Условием стабильности является $|a| < 1$. В зависимости от знака a растягивается низкочастотный ($a > 0$) или высокочастотный ($a < 0$) диапазон путем неравно-

мерного распределения точек преобразования на единичной окружности. Формально это может быть выражено следующим образом:

$$\hat{\omega} = \omega + 2\arctan\left(\frac{a \cdot \sin(\omega)}{1 - a \cdot \cos(\omega)}\right) \text{ для } \begin{cases} z = e^{j\omega} \\ \hat{z} = e^{j\hat{\omega}} \end{cases} \quad (7.33)$$

то есть осуществляется билинейное преобразование z -плоскости в новую искривленную \hat{z} -плоскость (рис. 7.15).

Таким образом, WDFT является обобщением ДПФ с частотными компонентами, расположенными неравномерно, но регулярно на единичной окружности. В матричной записи (с $\hat{X}(k)$, обозначаящим $\hat{X}(z_k)$) WDFT может быть представлено следующим выражением:

$$\begin{bmatrix} \hat{X}(0) \\ \hat{X}(1) \\ \vdots \\ \hat{X}(N-1) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & A(z_0) & \dots & A(z_0)^{N-1} \\ 1 & A(z_1) & \dots & A(z_1)^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & A(z_{N-1}) & \dots & A(z_{N-1})^{N-1} \end{bmatrix}}_{\mathbf{D}} \cdot \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix}. \quad (7.34)$$

Вычислительные аспекты прямого и обратного преобразования WDFT изложены в главе 5.

7.5.2 Аппроксимация шкалы барков

Первый шаг при использовании WDFT в психоакустической модели – проектирование соответствующего фазового преобразования. Частотные коэффициенты z -преобразования должны быть представлены равномерно в перцептуальной шкале. В работе [22] было показано, что фазовое звено первого порядка достаточно хорошо аппроксимирует перцептуальную шкалу барков, при этом значение коэффициента фазового фильтра для заданной частоты дискретизации определяется по следующему выражению:

$$a_{bark} = 0.1957 - 1.048 \cdot \left[\frac{2}{\pi} \arctan\left(0.07212 \frac{f_s}{1000}\right) \right]^2. \quad (7.35)$$

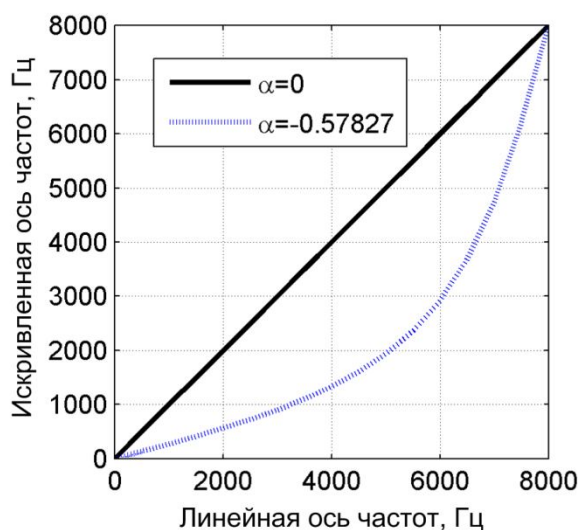


Рис. 7.16 – Искривление частотной оси в соответствии с критической шкалой Барков

Для случая $f_s = 16$ кГц коэффициент $a_{\text{барк}} = -0.57827$. На рис. 7.16 представлена трансформация (деформирование) частотной оси согласно шкале Барков.

Так как ДПФ имеет равномерное частотное разрешение, в то время как ширина критических частотных полос строго изменяется с их местоположением на частотной шкале, то различное количество коэффициентов преобразования ассоциируется с конкретной критической частотной полосой. В части А табл. 7.5 количество коэффициентов в группах варьируется от 3 до 38, в то время как для WDFT той же размерности не отдается предпочтения ни одной из полос, все коэффициенты преобразования распределены практически равномерно (часть Б). Более наглядно распределение коэффициентов преобразования по критическим частотным полосам показано на рис. 7.17.

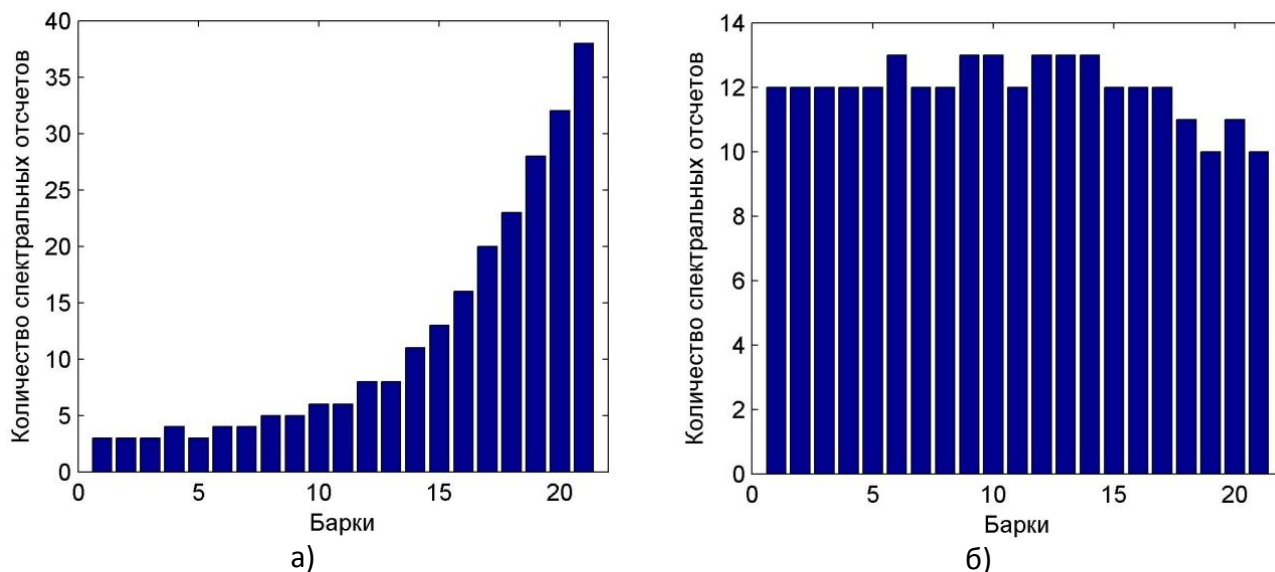


Рис. 7.17 – Распределение коэффициентов преобразования по критическим частотным полосам

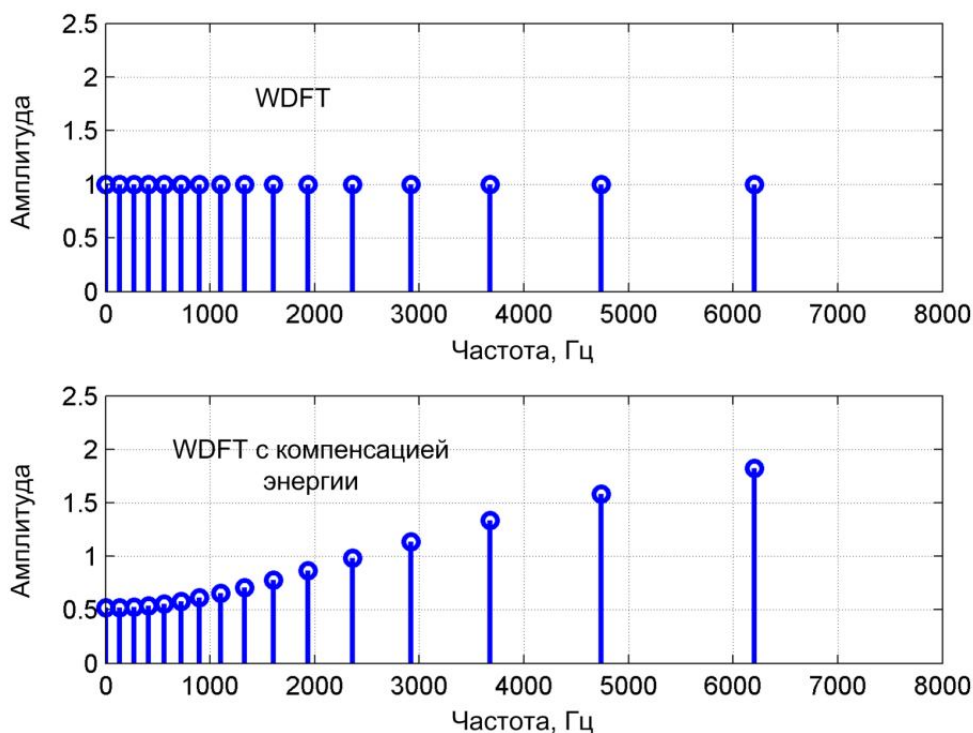


Рис. 7.18 – Компенсация энергии WDFT ($a_{\text{барк}} = -0.57827$) единичного импульса

WDFT в его оригинальной форме не сохраняет энергию сигнала в соответствующих частях единичной окружности до и после всепропускающего преобразования (рис. 7.18).

Так как искривление приводит к растяжению одного частотного диапазона при одновременном сжатии другого, оно должно сопровождаться масштабированием WDFT-амплитуд с целью коррекции уровней энергии в каждой критической частотной полосе. В работах [7,23] предлагается следующая коррекция:

$$\hat{X}_{PC}(z) = \frac{\sqrt{1-a^2}}{1-a \cdot z} \hat{X}(z). \quad (7.36)$$

Табл. 7.5 – Сравнение распределения коэффициентов ДПФ и WDFT в критических частотных полосах

Часть А (размер ДПФ = 512, $f_s = 16$ кГц)				Часть Б (размер WDFT = 512, $f_s = 16$ кГц)			
Крит. пол.	Диапазон коэф.	Кол-во	Диапазон частот, Гц	Крит. пол.	Диапазон коэф.	Кол-во	Диапазон частот, Гц
1	1 – 3	3	31 - 94	1	1 – 12	12	8 – 100
2	4 – 6	3	125 - 188	2	13 – 24	12	109 – 202
3	7 – 9	3	219 - 281	3	25 – 36	12	210 – 305
4	10 – 13	4	313 - 406	4	37 – 48	12	314 – 412
5	14 – 16	3	438 - 500	5	49 – 60	12	421 – 523
6	17 – 20	4	531 - 625	6	61 - 73	13	533 – 650
7	21 – 24	4	656 - 750	7	74 - 85	12	660 – 776
8	25 – 29	5	781 - 906	8	86 - 97	12	787 – 912
9	30 – 34	5	938 - 1063	9	98 - 110	13	923 – 1073
10	35 – 40	6	1094 - 1250	10	111 - 123	13	1086 - 1254
11	41 – 46	6	1281 - 1438	11	124 - 135	12	1269 - 1443
12	47 – 54	8	1469 - 1688	12	136 - 148	13	1460 - 1680
13	55 – 62	8	1719 - 1938	13	149 - 161	13	1700 - 1961
14	63 – 73	11	1969 - 2281	14	162 - 174	13	1985 - 2302
15	74 – 86	13	2313 - 2688	15	175 - 186	12	2331 - 2690
16	87 – 102	16	2719 - 3188	16	187 - 198	12	2726 - 3174
17	103 – 122	20	3219 - 3813	17	199 - 210	12	3220 - 3792
18	123 – 145	23	3844 - 4531	18	211 - 221	11	3851 - 4513
19	146 – 173	28	4563 - 5406	19	222 - 231	10	4588 - 5328
20	174 – 205	32	5438 - 6406	20	232 - 242	11	5419 - 6412
21	206 – 243	38	6438 - 7594	21	243 - 252	10	6520 - 7533
22	244 – 256	13	7625 - 8000	22	253 - 256	4	7650 - 8000

На примере широкополосного аудиосигнала, который содержит голосовое сообщение (первые 3.5 с) и музыкальный фрагмент с вокальным сопровождением (рис. 7.19 б) и в)) можно увидеть сравнение двух преобразований ДПФ и WDFT. Как видно из рис. 7.19, WDFT-решение дает достаточно четкий результат – оценка энергии более точная. Данное преобразование более детально показывает частотные характеристики сигнала в НЧ-области, кроме того ВЧ-область также имеет лучшее разрешение и представляется более детально. Представленные

факты свидетельствуют о том, что WDFТ может действительно служить базисом для психоакустической модели.

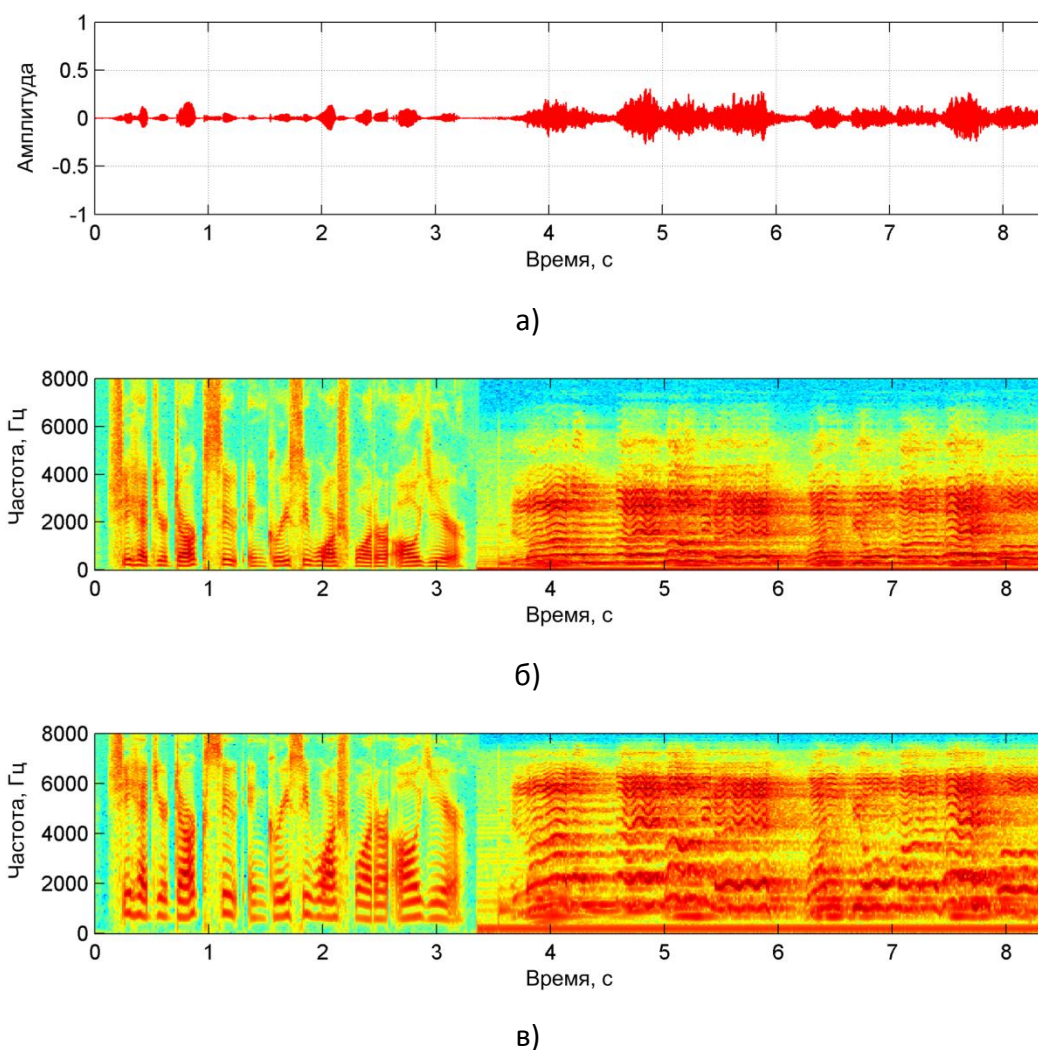


Рис. 7.19 – Широкополосный сигнал во временной области (а) и его спектрограммы, полученные при помощи ДПФ (б) и WDFТ с коэффициентом $a_{bark} = -0.57827$ (в)

7.5.3 Оценка порогов маскирования в WDFТ-области

В известной работе Джонстона [15] психоакустическая модель основана на ДПФ: расчет ДПФ взвешенного сегмента сигнала, объединение коэффициентов преобразования в группы, соответствующие критическим частотным полосам восприятия, и расчет энергии в критических частотных полосах. Достижение приемлемого спектрального разрешения в наименьших критических частотных полосах требует использования ДПФ с достаточно длинным временным окном. Поэтому концептуальная простота и эффективность нивелируются недостаточным временным разрешением, неприемлемым для анализа более тонкого феномена, такого как маскирование назад («pre-masking»).

Исследования, проведенные в работах [7,23] показали, что малоразмерное WDFТ, может успешно заменить ДПФ с большой длиной выборки. Это оказалось возможным благодаря тому, что WDFТ позволяет разместить частотные компоненты в соответствии с распределением кри-

тических частотных полос, поэтому в психоакустической модели на базе WDFT могут быть уравновешены как хорошее частотное, так и временное разрешение.

Для использования свойств WDFT процедура оценки порога маскирования на базе общей психоакустической модели Джонсона была модифицирована и подстроена под широкополосный речевой сигнал с полосой 8 кГц. Основные шаги оценки порога маскирования следующие: расчет энергии критических частотных полос по WDFT-спектру мощности; свертка с функцией распространения; вычитание относительного смещения порога; нормализация и сравнение с абсолютным порогом слышимости.

Энергия рассчитывается в критических частотных полосах путем суммирования компонент спектра мощности в каждой критической частотной полосе. Коэффициенты преобразования группируются в субполосы в соответствии с табл. 6, в каждой из которых осуществляется оценка порогов маскирования. Как упоминалось выше, в случае WDFT существует неравномерная концентрация энергии в окрестностях преобразуемых центральных частот. Поэтому, для вычисления энергии, сосредоточенной в критических частотных полосах, необходимо учесть масштабирование (7.36).

Дискретное представление энергии критической частотной полосы может быть выражено как взвешенная сумма компонент спектра мощности:

$$E_b = \frac{1}{2\pi} \sum_{k \in K_b} \Delta\omega_k |\hat{S}(\omega_k)|^2, \quad (7.37)$$

где b – номер критической частотной полосы; K_b – количество коэффициентов преобразования, попадающих в b -ю критическую частотную полосу; k – индекс коэффициента преобразования (спектрального отсчета); $\hat{S}(\omega_k)$ – оценка спектра оригинального речевого сигнала.

На следующем шаге вычисляется свертка энергий критических частотных полос E_b с функцией распространения по базилярной мембране SF_b для учета распространения маскирования:

$$C_b = E_b \cdot SF_b. \quad (7.38)$$

Относительное смещение порога в каждой критической частотной полосе рассчитывается с использованием меры тональности. Для определения типа маскера («шумоподобный» или «тональный») используется мера пологости спектра, которая определяется как отношение среднего геометрического μ_{gb} к среднему арифметическому μ_{ab} спектральных компонент внутри критической полосы b :

$$SFM_b = 10 \log_{10} \left(\frac{\mu_{gb}}{\mu_{ab}} \right), [\text{дБ}] \quad (7.39)$$

Данная оценка может быть затруднена в случае обычных моделей с малым разрешением, так как количества ДПФ-коэффициентов в низкочастотных полосах обычно не достаточно для эффективного расчета SFM . В подобных случаях SFM рассчитывается для всего спектра или predeterminedных тональностей для каждой критической частотной полосы. В случае WDFT, коэффициенты преобразования группируются равномерно в критических частотных полосах, следовательно, SFM может быть определена для каждой полосы отдельно.

Смещение порога маскирования O_b оценивается по следующему выражению:

$$O_b = ton_b \cdot (14.5 + b) + (1 - ton_b) \cdot 5.5, [\text{дБ}] \quad (7.40)$$

где ton_b – индекс тональности, определяемый как

$$ton_b = \min\left(\frac{SFM_b}{-60 \text{ дБ}}, 1\right). \quad (7.41)$$

Для получения энергии порога маскирования относительное смещение (7.40) вычитается из свернутого спектра критической полосы:

$$P_{TT,b} = 10^{(\log_{10} C_b - O_b/10)}, \quad (7.42)$$

а затем осуществляется нормализация и сравнение с абсолютным порогом слышимости.

7.5.4 Эмпирическое сравнение моделей аппроксимации критических частотных полос

Положим, что аудио сигнал, показанный на рис. 7.20 оцифрован с частотой дискретизации 32кГц и используется три реализации анализаторов энергии в критических частотных полосах на базе:

- процедуры вычисления WDFT: формат 256 отсчетов, фазовое звено первого порядка, $\alpha = -0.7056$, обработка сигнала ведется без перекрытия фреймов;
- алгоритма БПФ формата 1024 отсчетов с цепочкой фазовых звеньев первого порядка [24] и коэффициентом $\alpha = -0.7056$, обработка сигнала ведется с 50% перекрытием фреймов и временным окном Хэмминга;
- алгоритма БПФ формата 1024 отсчета, с 50% перекрытием фреймов и временным окном Хэмминга.

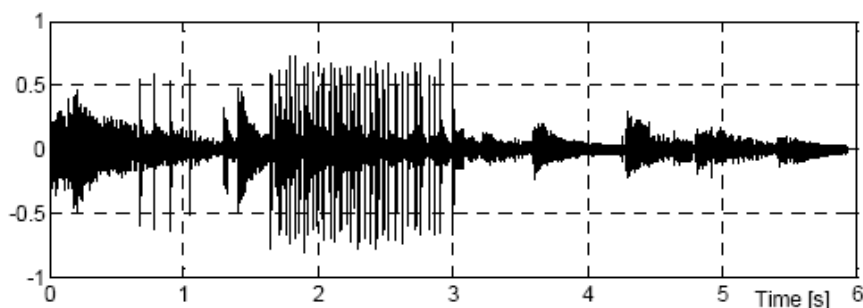
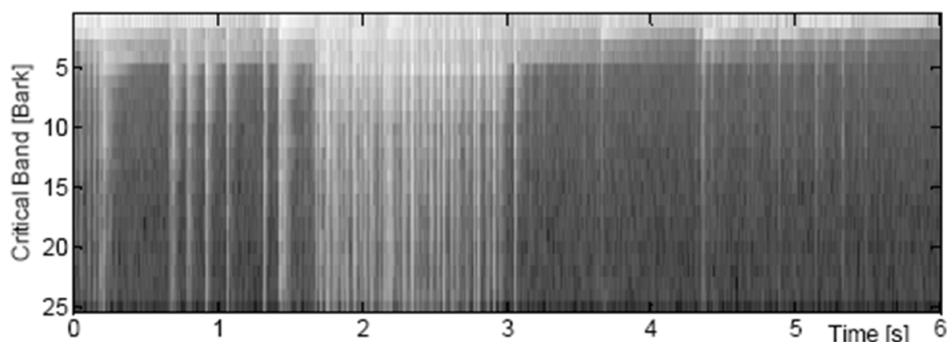


Рис. 7.20 – Тестовый аудиосигнал – кастанеты

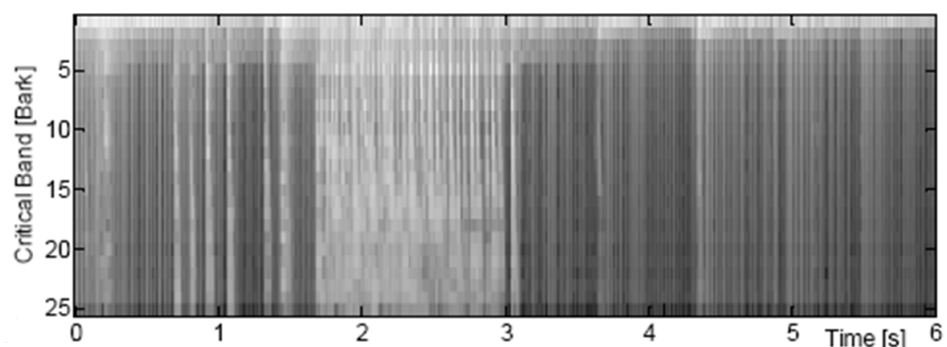
Сравнение трех данных анализаторов осуществляется на оценке СПМ в критических частотных полосах, которую можно увидеть на соответствующих спектрограммах (рис. 7.21). Короткий фрейм анализа сигнала на базе процедуры вычисления WDFT и алгоритма БПФ формата 1024 отсчетов с цепочкой фазовых звеньев первого порядка (рис. 7.21 а) и б) соответственно определяет разрешающую способность анализа и точность оценок СПМ: из спектрограмм видно, что в области низких частот (первых Барков) энергия локализована. Погрешности, обусловленные обработкой сигнала в цепочках фазовых фильтров, проявляются в виде шумовых артефактов (узкие линии на спектрограмме). Большой размер обрабатываемого фрейма сигнала в

случае анализатора СПМ в критических частотных полосах на основе алгоритма БПФ (рис. 7.21 в)) обуславливает размытие спектрограммы в области высоких частот (высокая разрешающая способность алгоритма – большое число коэффициентов преобразования попадает в соответствующие барки, в то время как в области низких частот – оценки СПМ осуществляются с большей статистической погрешностью).

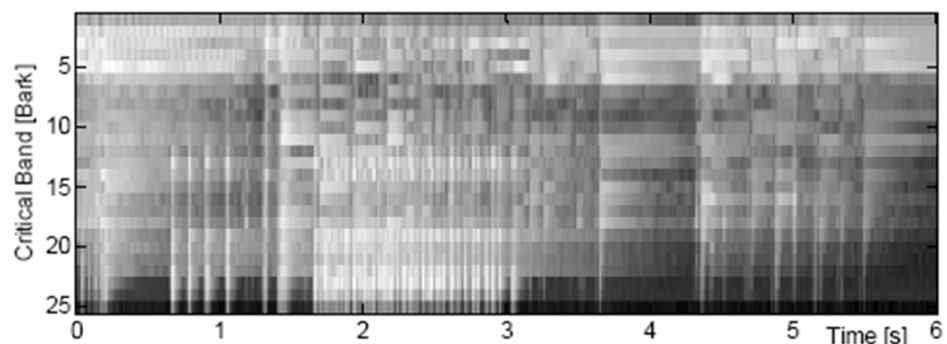
Таким образом, психоакустическая модель, построенная на базе WDFT, обеспечивает статистически равнозначный анализ энергии сигнала в критических частотных полосах – Барках.



а)



б)



в)

Рис. 7.21 – Аппроксимация шкалы Барков на основе : а) WDFT, б) БПФ с цепочкой фазовых звеньев первого порядка, в) БПФ

7.6 Список использованных источников

1. Spanias A., Painter T., Atti V., Audio signal processing and coding, Hoboken, NJ: John Wiley & Sons, Inc., 2007, 464 p.
2. Petrovsky Al., Krahe D., Petrovsky A.A., Real-time wavelet packet-based low bit rate audio coding on a dynamic reconfiguration system, Audio Engineering Society, Convention paper 5778, 114th Convention, 2003 March 22-25, Amsterdam, The Netherlands, 22 pages.
3. Петровский А., Субполосное перцептуальное кодирование звуковых и речевых сигналов, Мн: Бестпринт, 2008 г., 220 с.
4. Zwicker E., Fastl H., Psychoacoustics: Facts and Models, Berlin, Germany: Springer-Verlag, 1990, 380 p.
5. Vetterli M., Kovacevic J., Wavelets and Subband coding, Englewood Cliffs. NJ: Prentice-Hall, 1995, 488 p.
6. Petrovsky A., Parfieniuk M., Borowicz A., Warped DFT based perceptual noise reduction system, Audio Engineering Society, Convention Paper 6035, presented at the 116th Convention, 8-11 May 2004, Berlin, Germany, 16 pages.
7. Borowicz A., Parfieniuk M., Petrovsky A.A., An application of the warped discrete Fourier transform in the perceptual speech enhancement, Speech Communication, vol. 48, ELSEVIER, 2006, pp. 1024-1036.
8. Wickerhauser M.V. Adaptive Wavelet Analysis from Theory to Software, A.K. Peters Ltd., Massachusetts, 1994, 486 p.
9. Sinha D., Tewfik A.H., Low bit rate transparent audio compression using adapted wavelets, IEEE Trans. Signal Processing, vol. 41, no. 12, Dec. 1993, pp.3463-3479.
10. ITU-R Rec. BS.1387, Method for Objective Measurements of Perceived Audio Quality, 1998.
11. Petrovsky Al., Parfieniuk M., Borowicz A., Petrovsky A., Auditory modeling via frequency warped transforms, Audio Engineering Society, Convention paper 7446, 124th Convention, May 2008, Amsterdam, The Netherlands, 15 pages.
12. Carnero B., Drygajlo A., Perceptual speech coding and enhancement using frame-synchronized fast wavelet packet transform algorithms, IEEE Trans. Signal Processing, vol. 47, no. 6, June 1999, pp.1622-1635.
13. Ковалгин Ю.А., Вологдин Э.И., Цифровое кодирование звуковых сигналов, Санкт-Петербург: Корона-Принт, 2004 г., 240 с.
14. Петровский Ал.А., Построение психоакустической модели в области вейвлет-коэффициентов для перцептуальной обработки звуковых и речевых сигналов, Речевые технологии, № 4, 2008, с. 61–71.
15. Johnston J.D., Transform coding of audio signals using perceptual noise criteria, IEEE Transactions on Selected Areas Communication, February, 1988, vol. 6, pp. 314–323.

-
16. Petrovsky Al., Petrovsky A. A., Transitional speech segments modeling by matching pursuit with a dictionary based on the psychoacoustic adaptive WP, Proc. of the 11th Int. Conf. "Speech and Computer" (SPECOM'2006), 25-29 June, St. Petersburg, 2006, pp. 380-385.
 17. Парфенюк М., Петровский А.А., Параунитарные банки фильтров на основе алгебры кватернионов: теория и применение, Цифровая обработка сигналов, №1, Москва, 2008, с.22-36.
 18. Piotrowski A., Parfieniuk M., Cyfrowe banki filtrow: analiza, synteza I implementacja dla systemow multimedialnych, Wydawnictwo Politechniki Bialostockiej, Bialystok, 2006, 389s.
 19. Parfieniuk M., Petrovsky A., Inherently lossless structures for eight- and six-channel linear-phase paraunitary filter banks based on quaternion multipliers, Signal processing, vol. 90, issue 6, June 2010, pp. 1755-1767.
 20. Parfieniuk M., Petrovsky A., Quaternionic lattice structures for four-channel paraunitary filter banks, EURASIP Journal on Advances in Signal Processing, vol. 2007, No 9, Article ID 37481, 12 pages.
 21. Makur A., Mitra S.K., A Warped Discrete-Fourier Transform: Theory and Applications, IEEE Trans. Circuits Systems I, 2001., vol. 48, no. 9, pp. 1086–1093.
 22. Smith J.O., Abel J.S., Bark and ERB Bilinear Transforms, IEEE Transactions on Speech and Audio Processing, June 1999, vol. 7, pp. 697–708.
 23. Parfieniuk M., Petrovsky A., Warped DFT as the basis for psychoacoustical model, The proc. of the IEEE International conference an Acoustic, Speech, Signal processing, ICASSP-2004, Montreal, 2004, pp. IV-185-188
 24. Oppenheim A.V., Johnson D.H., Steiglitz K., Computation of spectra with unequal resolution using the FFT, Proceedings IEEE, February 1971, vol. 59, pp. 299–301.

8 ВЕКТОРНОЕ КВАНТОВАНИЕ И КОМПРЕССИЯ ЗВУКОВЫХ СИГНАЛОВ

Лившиц М.З., Павловец А.Н., Лихачев Д.С., Петровский А.А.

АННОТАЦИЯ

В главе дается определение векторного квантования и описываются его основные принципы. Представлены варианты построения векторных квантователей, описан алгоритм обучения (формирования) кодовых книг. Перечислены основные меры искажения, приведены примеры квантования коэффициентов линейного предсказания. Рассматривается метод векторного квантования с переменной размерностью векторов для параметров гармонической модели речевого сигнала. Особенностью метода является использование закономерностей психоакустики в процедуре квантования амплитуд, что позволяет повысить качество реконструированного речевого сигнала и снизить вычислительную сложность алгоритмов квантования.

Ключевые слова – векторное квантование, компрессия сигналов, квантование с расщеплением, многоуровневое квантование.

СОДЕРЖАНИЕ

8.1	Понятие векторного квантования	163
8.2	Меры искажений, используемые в векторном квантовании.....	167
8.3	Формирование кодовой книги.....	169
8.4	Векторное квантование с расщеплением	171
8.5	Многоуровневое векторное квантование.....	172
8.6	Векторное квантование коэффициентов линейного предсказания.....	172
8.6.1	Линейное предсказание речи.....	172
8.6.2	Векторное квантование LPC-параметров (LSF-коэффициентов).....	175
8.7	Квантование амплитуд гармоник речевого сигнала с использованием закономерностей психоакустики.....	180
8.7.1	Векторное квантование с переменной размерностью векторов.....	182
8.7.2	VDVQ с психоакустически обоснованным ограничением длины вектора	183
8.8	Список использованных источников.....	188

8.1 Понятие векторного квантования

Квантование – процесс аппроксимации непрерывных сигналов дискретными значениями, играющий важную роль при сжатии данных или кодировании с целью сокращения числа бит, необходимого для передачи или хранения сигналов с требуемой точностью [1]. Независимое квантование каждого значения сигнала или параметра называется скалярным квантованием (СК). Операция квантования здесь сводится к тому, что всем величинам квантуемого па-

раметра, попавшим в некоторый интервал, приписывается одно и то же заданное значение. Скалярное квантование бывает равномерным и неравномерным. В последнем случае меньшие интервалы используются в областях с большей вероятностью появления квантуемых величин, а большие интервалы – в областях с меньшей вероятностью появления.

Совместное квантование блока параметров называется блочным или векторным квантованием (ВК). Векторное квантование представляется как процесс исключения избыточности, в котором эффективно используются четыре взаимосвязанных свойства векторных параметров: линейная зависимость (корреляция), нелинейная зависимость, форма функции плотности вероятности (ФПВ) и многомерность вектора, тогда как при скалярном квантовании можно эффективно использовать только линейную зависимость и форму ФПВ. Нелинейная зависимость играет существенную роль при квантовании спектральных параметров речи, тогда как многомерность важна при квантовании формы сигнала. Вследствие относительно высокой стоимости векторного квантования (зависимость стоимости от размерности вектора и количества бит на координату обычно имеет экспоненциальный характер) на современном уровне развития вычислительной техники преимущества векторного квантования проявляются в основном при скоростях передачи до 1 бита на параметр. Как раз в этой области рабочие характеристики скалярных квантователей резко ухудшаются. Таким образом, к числу основных преимуществ данного метода кодирования следует отнести следующие характеристики: а) возможность устранить корреляционные связи между координатами вектора признаков; б) возможность учесть нелинейные зависимости между координатами вектора признаков. Все это позволяет существенно снизить избыточность цифрового представления сигнала, а следовательно, сократить алфавит возможных состояний акустической модели.

При использовании векторного квантования предполагается, что $x = [x_1, x_2, \dots, x_N]^T$ представляет собой N -мерный вектор, компоненты которого $\{x_k, 1 \leq k \leq N\}$ – действительные случайные величины с непрерывным распределением амплитудных значений (T - символ транспонирования). При ВК вектор x отображается в N -мерный действительный вектор y с дискретными значениями амплитуд. Этот процесс называется квантованием x в y , а y представляет собой квантованное значение x . Процесс квантования можно записать в виде следующего выражения:

$$y = q(x), \quad (8.1)$$

где $q(\cdot)$ – оператор квантования, y – преобразованный вектор (выходной вектор), соответствующий x . Обычно y принимает одно значение из ограниченного множества $Y = \{y_i, 1 \leq i \leq L\}$, где $y_i = [y_{i1}, y_{i2}, \dots, y_{iN}]^T$. Множество Y называется кодовой книгой преобразования или просто кодовой книгой. L – размер кодовой книги, а $\{y_i\}$ – множество кодовых векторов. Размер кодовой книги называют также числом уровней (этот термин пришел из терминологии скалярного квантования). Таким образом, можно говорить о кодовой книге с L уровнями или об L -уровневом квантователе. Для построения такой кодовой книги N -мерное пространство случайного вектора x разделяется на L областей или ячеек $\{C_i, 1 \leq i \leq L\}$ и с каждой ячейкой C_i связывается вектор y_i . Квантователь назначает кодовый вектор y_i , если x лежит в C_i :

$$q(x) = y_i, \text{ если } x \in C_i. \quad (8.2)$$

Процесс построения кодовой книги известен также как процесс обучения или заполнения кодовой книги. На рис. 8.1 приведен пример разделения двумерного пространства ($N = 2$) при векторном квантовании.

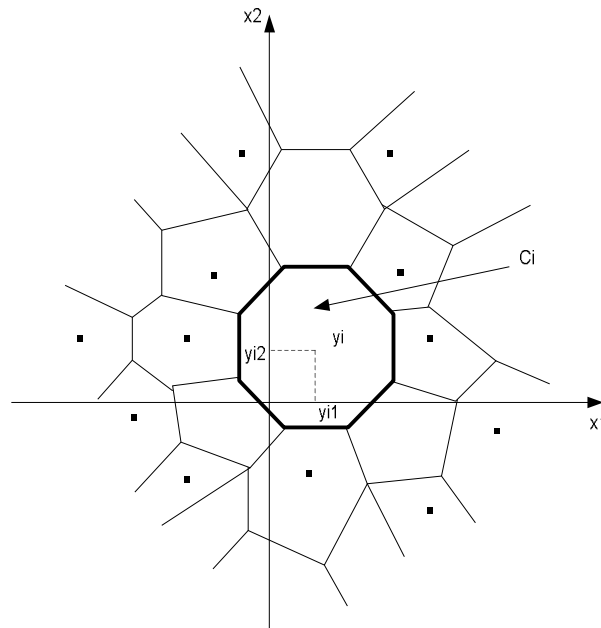


Рис. 8.1 – Разделение двумерного пространства ($N = 2$) на $L = 24$ ячейки (формы разных ячеек могут отличаться весьма значительно)

Область, окруженная утолщенными линиями, представляет собой ячейку C_i . Любой входной вектор x , лежащий в ячейке C_i , квантуется как u_i . Положения кодовых векторов других ячеек обозначены точками. Общее число кодовых векторов в приведенном на примере составляет $L = 24$. При $N = 1$ векторное квантование вырождается в скалярное квантование. На рис. 8.2 показан пример разделения действительной оси при скалярном квантовании.

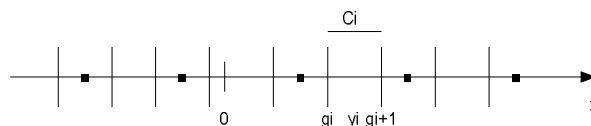


Рис. 8.2 – Разбиение действительной прямой на $L = 10$ ячеек (интервалов) при скалярном квантовании

Значения кода (выходные, или преобразованные уровни) отмечены точками. Здесь также любое значение входной величины x , которое лежит в интервале C_i , квантуется как u_i (число уровней квантования $L = 10$). Скалярное квантование имеет особенность, состоящую в том, что хотя размеры ячеек могут быть различными, все они имеют одинаковую форму, а именно все они представляются интервалами на действительной оси. Для сравнения на рис. 8.1 были показаны двумерные ячейки, имеющие различную форму.

Благодаря возможности построения в многомерном пространстве ячеек различной формы векторное квантование имеет преимущества перед скалярным квантованием.

При квантовании x в y возникает ошибка квантования. Отклонение x от y может быть определено мерой искажения $d(x, y)$, называемой также мерой расхождения или мерой расстояния.

В случае передачи векторов $y(n)$ в различные моменты времени n можно определить результирующее среднее искажение по следующему выражению [1]:

$$D = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{n=1}^M d[x(n), y(n)]. \quad (8.3)$$

Если векторный процесс $x(n)$ обладает свойствами стационарности и эргодичности (эргодичность позволяет заменить среднее по ансамблю на среднее по отсчетам (или по времени)), то выборочное среднее в пределе стремится к математическому ожиданию [1]:

$$\begin{aligned} D = E[d(x, y)] &= \sum_{i=1}^L P(x \in C_i) E[d(x, y_i) | x \in C_i] = \\ &= \sum_{i=1}^L P(x \in C_i) \int_{x \in C_i} d(x, y_i) p(x) dx, \end{aligned} \quad (8.4)$$

где $E[\cdot]$ – математическое ожидание, $P(x \in C_i)$ – вероятность того, что x находится в C_i , $p(x)$ – многомерная функция плотности вероятности (ФПВ) величины x .

При передаче каждый вектор y_i кодируется двоичным кодовым словом C_i , содержащим B_i бит. В общем случае различные кодовые слова имеют различную длину. При этом скорость передачи T определяется выражениями [1]:

$$T = B \cdot F_c \frac{\text{бит}}{\text{с}}, \quad B = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{n=1}^M B(n) \frac{\text{бит}}{\text{вектор}}, \quad (8.5)$$

где $B(n)$ – число бит, использованных для кодирования вектора $x(n)$ в момент времени n , F_c – число кодовых слов, передаваемых за одну секунду.

Полезно также определить среднее число бит на параметр или координату вектора как отношение числа бит, использованных для кодирования вектора, к количеству координат в векторе – N :

$$R = \frac{B}{N} \text{ бит/координата}. \quad (8.6)$$

При использовании кодовой книги размером L максимальное число бит, необходимое для кодирования каждого вектора, можно определить согласно следующему выражению:

$$B_{max} = \log_2 L. \quad (8.7)$$

Процесс векторного квантования осуществляется следующим образом. Кодер и декодер содержат кодовую книгу, состоящую из векторов (блоков) параметров C_i , представляющих входные векторы параметров x . Для каждого блока входных параметров в кодере реализуется поиск записи C_i , обеспечивающей наилучшее соответствие входному вектору x . Индекс кодового вектора, наилучшим образом представляющего блок входных параметров, передается в

декодер, который использует кодовую книгу как таблицу для реконструкции входного сигнала. На рис. 8.3 представлена наиболее общая схема векторного квантования. При построении системы сжатия данных квантователь должен проектироваться таким образом, чтобы для заданной скорости передачи искажения выходного сигнала были минимальны. При проектировании квантователя один из главных вопросов заключается в выборе меры искажений.

8.2 Меры искажений, используемые в векторном квантовании

Мера искажений найдет применение лишь в том случае, если она выражается в удобной математической форме, пригодной для анализа и расчетов, и если она отражает свойства субъективного восприятия, так что различия в величинах искажений характеризуют соответствующие различия качества речи.

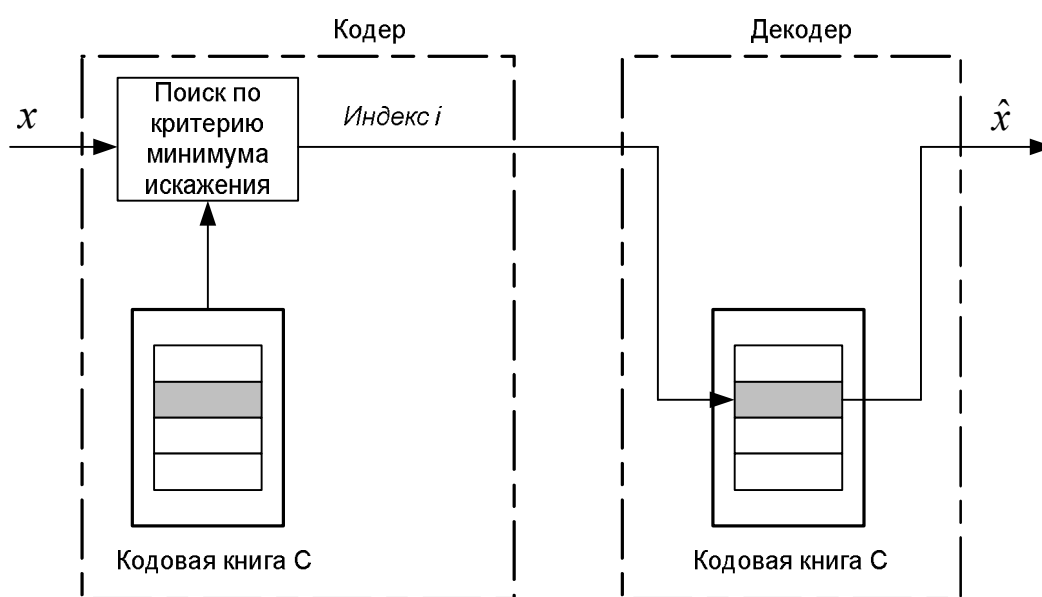


Рис. 8.3 – Упрощенная схема процесса векторного квантования

Большинство из используемых в настоящее время мер искажений поддается математическому анализу и отражает некоторые особенности восприятия. Тем не менее в одних случаях уменьшение искажения на несколько децибел весьма заметно на слух, а в других случаях может быть и вовсе незаметно. Несмотря на то, что объективные меры искажений могут служить необходимыми и полезными инструментами при проектировании систем кодирования речи, для принятия правильных решений о выборе путей улучшения характеристик системы необходимы периодические субъективные испытания качества. Рассмотрим наиболее распространенные меры искажений.

1. Среднеквадратичная ошибка или отклонение (СКО) определяется по формуле [1-3]:

$$d_2(x, y) = \frac{1}{N} (x - y)^T (x - y) = \frac{1}{N} \sum_{k=1}^N (x_k - y_k)^2, \quad (8.8)$$

где x, y – анализируемые векторы, N – размерность векторов, x_k, y_k – k -е координаты векторов x, y соответственно.

2. Общая мера искажений основана на критерии L_r , определяемом в виде выражения [1]

$$d_r(x, y) = \frac{1}{N} \sum_{k=1}^N |x_k - y_k|^r. \quad (8.9)$$

Заметим, что при $r = 2$ общая мера искажения преобразуется в СКО. Часто используются два других значения $r = 1$ и $r = \infty$. Соответственно d_1 представляет собой среднее значение абсолютной ошибки, а d_∞ характеризует максимальную ошибку. При кодировании речи наибольшее распространение получила мера d_2 , хотя иногда используются меры d_1 и d_∞ .

3. Взвешенная среднеквадратичная ошибка. При использовании меры d_2 предполагается, что искажения, обусловленные квантованием различных параметров $\{x_k\}$, учитываются с одинаковым весом. В общем случае можно ввести неравные веса и тем самым увеличить вклад искажений некоторых параметров в общую меру искажений. Взвешенную СКО можно определить как

$$d_w(x, y) = (x - y)^T W (x - y), \quad (8.10)$$

где W – положительно определенная взвешивающая матрица.

Часто в задачах распознавания образов принимают $W = \Gamma^{-1}$, где Γ – ковариационная матрица случайного вектора x :

$$\Gamma = E[(x - \bar{x})(x - \bar{x})^T], \quad (8.11)$$

где \bar{x} – среднее по ансамблю реализаций (в нашем случае математическое ожидание). В этом случае d_w можно представить в виде следующего выражения, которое в данном случае известно под названием расстояние Махаланобиса:

$$d_w(x, y) = (x - y)^T \Gamma^{-1} (x - y). \quad (8.12)$$

Если взвешивающая матрица W не только положительно определена, но и симметрична (как в случае расстояния Махаланобиса), ее можно представить в виде произведения

$$W = P^T P, \quad (8.13)$$

где P – нижняя треугольная матрица с единичной главной диагональю. Векторы x и y можно преобразовать в новый набор векторов $\tilde{x} = Px$ и $\tilde{y} = Py$, тогда

$$d_w(x, y) = (Px - Py)^T (Px - Py) = (\tilde{x} - \tilde{y})^T (\tilde{x} - \tilde{y}) = d_2(\tilde{x}, \tilde{y}). \quad (8.14)$$

Таким образом, взвешенная СКО исходных векторов равна СКО преобразованных векторов. Поэтому с целью упрощения вычислений перед векторным квантованием целесообразно произвести преобразование всех данных согласно выражению, представленному выше.

8.3 Формирование кодовой книги

Как упоминалось выше, для построения кодовой книги с L уровнями N -мерное пространство разделяется на L ячеек $\{C_i, 1 \leq i \leq L\}$ и каждой ячейке C_i приписывается вектор u_i . Тогда в квантователе назначается кодовый вектор y_i , если x лежит в ячейке C_i .

Квантователь называется оптимальным (обеспечивающим минимальные искажения), если принятая мера искажения минимизирована по всем L -уровневым квантователям. Существуют два необходимых условия оптимальности.

Первое условие заключается в том, что в оптимальном квантователе используется правило выбора по минимуму искажений, т.е. производится выбор ближайшей ячейки $q(x) = y_i$ тогда и только тогда, когда выполняется условие [1]

$$d(x, y_i) \leq d(x, y_j), \quad j \neq i, \quad 1 \leq j \leq L. \quad (8.15)$$

Другими словами, в квантователе осуществляется выбор кодового вектора, обеспечивающего минимальные искажения x . (При равенстве искажений на разных векторах назначается некоторое правило, позволяющее разрешить эту неопределенность и осуществить выбор.)

Второе необходимое условие оптимальности состоит в том, что каждый кодовый вектор y_i выбирается из условия минимизации среднего искажения в ячейке C_i . Таким образом, y_i есть вектор u_i , минимизирующий выражение:

$$D_i = E[d(x, y) | x \in C_i] = \int_{x \in C_i} d(x, y) p(x) dx. \quad (8.16)$$

Такой вектор называется центроидом ячейки C_i и записывается в виде

$$y_i = \text{cent}(C_i). \quad (8.17)$$

Для расчета центроида необходимо задать меру искажения. (Ячейки, определяемые таким образом, известны как ячейки ближайших соседей, ячейки Вороного или области Дирихле).

На практике задается набор обучающих векторов $\{x(n), 1 \leq n \leq M\}$. Некоторое подмножество M_i этих векторов попадает в ячейку C_i . Среднее искажение D_i определяется по следующей формуле:

$$D_i = \frac{1}{M_i} \sum_{x \in C_i} d(x, y_i). \quad (8.18)$$

Для критерия СКО или взвешенной СКО можно показать, что D_i минимизируется при определении центроида:

$$y_i = \frac{1}{M_i} \sum_{x \in C_i} x(n), \quad (8.19)$$

т.е. y_i представляет собой выборочное среднее обучающих векторов, содержащихся в ячейке C_i .

Поскольку решение задачи оптимального квантования в виде конечного выражения неизвестно, данная проблема решается путем итеративного улучшения заданного векторного квантователя. Однако, прежде чем улучшать квантователь, следует его инициализировать. Существует несколько примерно равноценных методов инициализации, рассмотрим один из них [1-3].

Шаг 1. Рассчитать центроид всего тренировочного множества.

Шаг 2. Рассчитать невзвешенное евклидово расстояние между центроидом и каждым вектором тренировочного множества.

Шаг 3. Выбрать вектор с максимальным расстоянием в качестве эталона.

Шаг 4. Рассчитать невзвешенное евклидово расстояние между эталонным вектором и каждым вектором тренировочного множества.

Шаг 5. Пусть M/L – отношение размера тренировочного множества к размерности кодовой книги. Найти M/L векторов, ближайших к эталонному вектору и рассчитать центроид (вектор начальной кодовой книги) для этой группы векторов.

Шаг 6. Уменьшить тренировочное множество путем удаления из него группы векторов, найденных на предыдущем шаге.

Шаг 7. Повторять шаги со 2-го по 6-й до тех пор, пока в тренировочном множестве не останется векторов.

Рассмотрим один из наиболее распространенных методов построения (тренировки) кодовой книги – алгоритм K -средних (его разновидностью является обобщенный алгоритм Ллойда (GLA) или алгоритм Линде-Бьюзо-Грэя (LBG)) [1,3,4]. В нашем случае $K = L$. Алгоритм разделяет набор обучающих векторов $\{x(n), 1 \leq n \leq M\}$ на L кластеров (ячеек) C_i таким образом, что удовлетворяются два необходимых условия оптимальности (8.15) и (8.16). Если обозначить через m номер итерации, а через $C_i(m)$ – i -й кластер на m -й итерации с центроидом $c_i(m)$, то алгоритм K -средних можно записать в следующем виде:

Шаг 1. Задание начальных значений. Положим $m = 0$. Выберем тем или иным подходящим методом набор начальных кодовых векторов $c_i(0)$, $1 \leq i \leq L$.

Шаг 2. Классификация. Классифицируем набор обучающих векторов $\{x(n), 1 \leq n \leq M\}$ по кластерам C_i с помощью правила “ближайшего соседа”: $x \in C_i(m)$ тогда и только тогда, когда $d[x, c_i(m)] \leq d[x, c_j(m)]$ для всех $j \neq i$.

Шаг 3. Коррекция кодового вектора. $m \leftarrow m + 1$. Произвести коррекцию кодовых векторов всех кластеров путем вычисления центроидов обучающих векторов каждого кластера $c_i(m) = \text{cent}(C_i(m))$, $1 \leq i \leq L$.

Шаг 4. Проверка на окончание процедуры. Если уменьшение величины общего искажения $D(m)$ на итерации m относительно $D(m - 1)$ меньше некоторого порога, процедура заканчивается. В противном случае переход на шаг 2.

Введем понятие тренировочного отношения, которое определяется как отношение размера тренировочной базы данных (обучающего множества векторов) к размеру кодовой книги. Для построения качественной кодовой книги тренировочное отношение должно быть в пределах от 10 до 50. До сих пор мы рассматривали построение кодовых книг без внутренней

структуры. Использование их в настоящее время невыгодно, т.к. ведет к длительным процессам тренировки и поиска в них либо к дополнительным затратам памяти.

Структурированные кодовые книги могут использовать быстрые методы поиска и занимают значительно меньший объем. В настоящее время разработано большое количество вариантов структурирования кодовых книг: квантование с расщеплением вектора (Split VQ – Split Vector Quantization), многоуровневое векторное квантование (MSVQ – Multistage Vector Quantization), векторное квантование с древовидной структурой (TSVQ – Tree-Structured Vector Quantization), решетчатое квантование (lattice quantization) и т.д. [3]. Рассмотрим подробнее структурированные кодовые книги на примере подходов SVQ и MSVQ.

8.4 Векторное квантование с расщеплением

В ВК с расщеплением входной вектор $x = [x_1, x_2, \dots, x_p]^T \in R^p$ расщепляется или делится на R субвекторов меньшей размерности $x = [x^{(1)} x^{(2)} x^{(3)} \dots x^{(R)}]^T$. Таким образом, i -й субвектор $x^{(i)}$ имеет размерность d_i , при этом $p = d_1 + d_2 + \dots + d_R$.

$$x^{(1)} = [x_1 x_2 \dots x_{d_1}]^T;$$

$$x^{(2)} = [x_{d_1+1} x_{d_1+2} \dots x_{d_1+d_2}]^T; \quad (8.20)$$

$$x^{(3)} = [x_{d_1+d_2+1} x_{d_1+d_2+2} \dots x_{d_1+d_2+d_3}]^T.$$

То есть, имеем R квантователей, по одному на каждый субвектор. Субвекторы $x^{(i)}$ индивидуально квантуются в $y_k^{(i)}$, при этом входной вектор x квантуется в $y = [y_k^{(1)} y_k^{(2)} y_k^{(3)} \dots y_k^{(R)}]^T \in R^p$. Квантователи проектируются (обучаются) путем использования соответствующих субвекторов в обучающем множестве. Частный случай ВК с расщеплением: $R = p$, $d_1 = d_2 = \dots = d_p = 1$ – получается скалярный квантователь. Предположим размерность $p = 10$. Полный 30-битный векторный квантователь будет иметь кодовую книгу из 2^{30} кодовых слов. В эквивалентном векторном квантовании с расщеплением из $R = 3$ расщеплений (частей) используются субвекторы с размерностями $d_1 = 3$, $d_2 = 3$, $d_3 = 4$, с каждым субвектором ассоциируется 10-битная кодовая книга, имеющая 2^{10} кодовых слов. Расщепленное ВК возможно, если

$$d(x, y) = \sum_{i=1}^R d(x^{(i)}, y^{(i)}). \quad (8.21)$$

Данное свойство истинно для L_r дистанции и для взвешенной L_2 дистанции, если матрица весов W является диагональной.

8.5 Многоуровневое векторное квантование

В многоуровневом ВК, включающем R уровней, используются R квантователей Q_1, Q_2, \dots, Q_R . Соответствующие кодовые книги обозначаются как C_1, C_2, \dots, C_R . Размеры этих кодовых книг – N_1, N_2, \dots, N_R . Суммарный размер кодовых книг определяется как $N = N_1 + N_2 + \dots + N_R$. Записи i -й кодовой книги $C_i - y_1^{(i)}, y_2^{(i)}, \dots, y_N^{(i)}$. Рис. 8.4 показывает структуру всей системы.

Процедура многоуровневого квантования следующая. Входной вектор x сначала квантуется при помощи квантователя Q_1 в $y_k^{(1)}$. Ошибка квантования $e_1 = x - y_k^{(1)}$ далее квантуется Q_2 в $y_k^{(2)}$. Ошибка второго уровня определяется как $e_2 = e_1 - y_k^{(2)}$ и квантуется на третьем уровне. Процесс повторяется, и на R -м уровне ошибка e_{R-1} квантуется Q_R в $y_k^{(R)}$, в результате на выходе схемы имеем ошибку e_R .

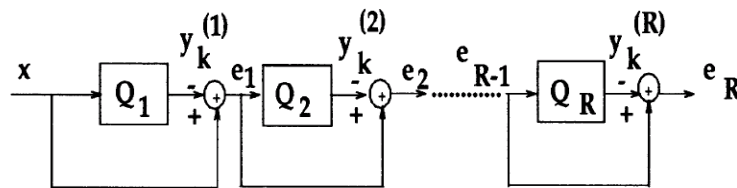


Рис. 8.4 – Схема многоуровневого векторного квантования

Таким образом, оригинальный вектор x квантуется в $y = y_k^{(1)} + y_k^{(2)} + \dots + y_k^{(R)}$. Общая ошибка квантования определяется как $x - y = e_R$. Уменьшение требуемой памяти и сложности поиска проиллюстрируем на простом примере. Полный (одинарный) 30-битный векторный квантователь будет иметь одну кодовую книгу 2^{30} кодовых векторов (не может использоваться на практике из-за большого объема требуемой памяти). Эквивалентный многоуровневый ВК с $R = 3$ уровнями будет иметь три 10-битные кодовые книги C_1, C_2, C_3 . Общее количество векторов, хранящихся в памяти, будет составлять $3 \times 2^{10} = 3072$, что практически реализуемо. Следовательно, сложность поиска также существенно снизится.

Обучение происходит, как было описано выше, только набор ошибок предыдущего уровня используется в качестве обучающего множества последующего уровня.

8.6 Векторное квантование коэффициентов линейного предсказания

8.6.1 Линейное предсказание речи

Линейное предсказание (LPC – Linear Predictive Coding) является одним из наиболее эффективных методов анализа речевого сигнала и применяется во многих кодерах речи. Этот метод становится доминирующим при сокращённом представлении речи с целью её низкоскоростной передачи и экономного хранения. Важность метода обусловлена высокой точностью получаемых оценок и относительной простотой вычислений.

Основной принцип метода линейного предсказания состоит в том, что текущий отсчёт речевого сигнала можно аппроксимировать линейной комбинацией предшествующих отсчё-

тов. Коэффициенты предсказания при этом определяются однозначно минимизацией среднего квадрата разности между отсчётами речевого сигнала и их предсказанными значениями (на конечном интервале). Коэффициенты предсказания – это весовые коэффициенты, используемые в линейной комбинации.

Основные положения метода линейного предсказания хорошо согласуются с моделью речеобразования, в которой речевой сигнал представлен в виде сигнала на выходе линейной системы с переменными во времени параметрами, возбуждаемой квазипериодическими импульсами (в пределах вокализованного сегмента) или случайным шумом (на невокализованном сегменте). Метод линейного предсказания позволяет точно и надёжно оценить параметры этой линейной системы с переменными коэффициентами.

Общий спектр, обусловленный излучением, речевым трактом и возбуждением, описывается с помощью линейной системы с переменными параметрами и передаточной функцией, выраженной следующей формулой [3,4]:

$$H(z) = \frac{G}{A(z)} = \frac{G}{1 - \sum_{k=1}^p \alpha_k z^{-k}}, \quad (8.22)$$

где $\{\alpha_k\}$ – множество коэффициентов, которые называются LPC-параметрами, или коэффициентами предсказателя; p – порядком предсказателя (число коэффициентов); G – коэффициент усиления.

Основная задача анализа на основе линейного предсказания заключается в непосредственном определении параметров $\{\alpha_k\}$ по речевому сигналу с целью получения хороших оценок его спектральных свойств путём использования уравнения (8.22). Вследствие изменения свойств речевого сигнала во времени коэффициенты линейного предсказания должны оцениваться на коротких сегментах речи. Основным подходом является определение параметров линейного предсказания таким образом, чтобы минимизировать дисперсию погрешности на коротком сегменте сигнала.

Как уже было сказано выше, очередной отсчет речевого сигнала может быть аппроксимирован линейной комбинацией предыдущих отсчетов, т.е.

$$\tilde{s}(n) = \sum_{k=1}^p \alpha_k s(n-k), \quad (8.23)$$

где $\tilde{s}(n)$ – предсказанное значение речевого отсчета, $s(n)$ – оригинальный отсчет речевого сигнала. Таким образом, ошибку предсказания можно представить как разность между оригинальным значением и предсказанным

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^p \alpha_k s(n-k). \quad (8.24)$$

Взяв z -преобразование от выражения (8.24), мы получим

$$E(z) = S(z)A(z), \quad (8.25)$$

где $A(z)$ – это инверсия от $H(z)$ в выражении (8.22), т.е. $A(z)$ – *инверсный фильтр*, уравнение которого может быть записано в следующем виде:

$$A(z) = 1 - \sum_{k=1}^p \alpha_k z^{-k}. \quad (8.26)$$

Из-за времязависимой природы речевого сигнала коэффициенты предсказания должны определяться на коротких сегментах речи (10 – 30 мс).

Таким образом, необходимо определить множество коэффициентов предсказателя, которые бы минимизировали ошибку на всем сегменте сигнала. Полученные параметры будут являться параметрами системной функции $H(z)$ модели речеобразования в выражении (8.22).

Средняя ошибка кратковременного предсказания определяется по следующему выражению:

$$E = \sum_n e^2(n) = \sum_n \left[s(n) - \sum_{k=1}^p \alpha_k s(n-k) \right]^2. \quad (8.27)$$

Для определения значений $\{\alpha_k\}$, минимизирующих ошибку E , необходимо взять частную производную по всем коэффициентам и приравнять ее к нулю

$$\frac{\partial E}{\partial a_i} = -2 \sum_n \left\{ \left[s(n) - \sum_{k=1}^p \alpha_k s(n-k) \right] s(n-i) \right\} = 0, \quad (8.28)$$

что дает следующее уравнение:

$$\sum_n s(n)s(n-i) = \sum_n \sum_{k=1}^p \alpha_k s(n-k)s(n-i). \quad (8.29)$$

Если изменить порядок суммирования в правой части уравнения (8.29), то получим следующее выражение:

$$\sum_n s(n)s(n-i) = \sum_{k=1}^p \alpha_k \sum_n s(n-k)s(n-i), \quad i = 1, \dots, p. \quad (8.30)$$

Если сделать замену:

$$\phi(i, k) = \sum_n s(n-i)s(n-k), \quad (8.31)$$

то уравнение (8.30) может быть записано в следующем виде:

$$\sum_{k=1}^p \alpha_k \phi(i, k) = \phi(i, 0), \quad i = 1, \dots, p. \quad (8.32)$$

Эта система из p уравнений с p неизвестными может быть эффективно решена относительно неизвестных коэффициентов $\{\alpha_k\}$.

В настоящее время применяется множество методов LP-анализа, но наибольшее распространение получили методы на основе автокорреляции и автоковариации, при этом первый

имеет меньшую вычислительную сложность и всегда обеспечивает синтез стабильного фильтра-предсказателя.

8.6.2 Векторное квантование LPC-параметров (LSF-коэффициентов)

Для речевых сегментов с высокой частотой основного тона LP-анализ получает фильтр-синтезатор с узкими областями спектрального резонанса. Для решения этой проблемы применяется расширение частотной полосы, которое позволяет уменьшить остроту этих пиков, а также решает некоторые проблемы числовой точности, связанные с близостью полюсов к единичной окружности. Количественно величина расширения (в герцах) определяется следующим образом [2,3]:

$$\Delta B = -\frac{1}{\pi F_S} \ln(\gamma), \quad (8.33)$$

где F_S – частота дискретизации (Гц).

Далее проводится модификация LPC-коэффициентов:

$$a'_k = a_k \gamma^k, \quad 1 \leq k \leq p. \quad (8.34)$$

Квантование LPC-параметров является одним из важнейших аспектов LP-анализа, поскольку минимизация ёмкости кодирования является основной целью в приложениях обработки речи. Хотя принципиально можно и непосредственно квантовать параметры предсказания, такой подход требует высокой точности представления (8 – 10 бит на параметр). Это связано с тем, что малые изменения параметров предсказания приводят к большим изменениям в расположении полюсов и, следовательно, к возможной неустойчивости синтезирующего LP-фильтра. Поэтому непосредственное квантование параметров не нашло применения.

Для квантования обычно используют альтернативное представление параметров линейного предсказания – линейные спектральные пары (частоты) (LSP – Line Spectral Pairs или LSF – Line Spectral Frequencies) [2,3].

Чтобы получить LSF-коэффициенты, p нулей функции $A_p(z)$ отражаются на единичную окружность посредством двух z -преобразований $P(z)$ и $Q(z)$ ($p + 1$)-го порядка:

$$P_{p+1}(z) = A_p(z) + z^{-(p+1)} A_p(z^{-1}); \quad (8.35)$$

$$Q_{p+1}(z) = A_p(z) - z^{-(p+1)} A_p(z^{-1}). \quad (8.36)$$

Из этого следует, что

$$A_p(z) = \frac{1}{2} [P_{p+1}(z) + Q_{p+1}(z)]. \quad (8.37)$$

LSF-коэффициенты представляют собой угловые позиции корней $P(z)$ и $Q(z)$ на единичной окружности в диапазоне $0 \leq \omega_i \leq \pi$. Они имеют следующие свойства: все корни $P(z)$ и $Q(z)$ лежат на единичной окружности; корни чередуются на единичной окружности, т.е. выполняется следующее неравенство:

$$0 \leq \omega_{q,0} \leq \omega_{p,0} \leq \omega_{q,1} \leq \omega_{p,1} \dots \leq \pi. \quad (8.38)$$

При выполнении данных условий синтезирующий фильтр $H(z)$ является стабильным. LSF обладают следующими важными свойствами:

- расстояние между LSF-коэффициентами определяет амплитуду спектральной плотности мощности (смотри рис. 8.5);
- блок из двух или трех близко расположенных LSF сигнализирует о наличии максимума в спектре мощности (соответствует формантной частоте), в то время как расположенные с большим промежутком LSF соответствуют минимуму (см. рис. 8.5);
- в общем случае спектральная чувствительность каждого LSF локализована, т.е. при небольшом изменении одного из LSF, спектр изменится только в окрестности этого LSF-параметра рис. 8.6.

Итак, для получения параметров LSF следует провести LP-анализ над взвешенными сегментами речевого сигнала и далее трансформировать полученные LPC-коэффициенты.

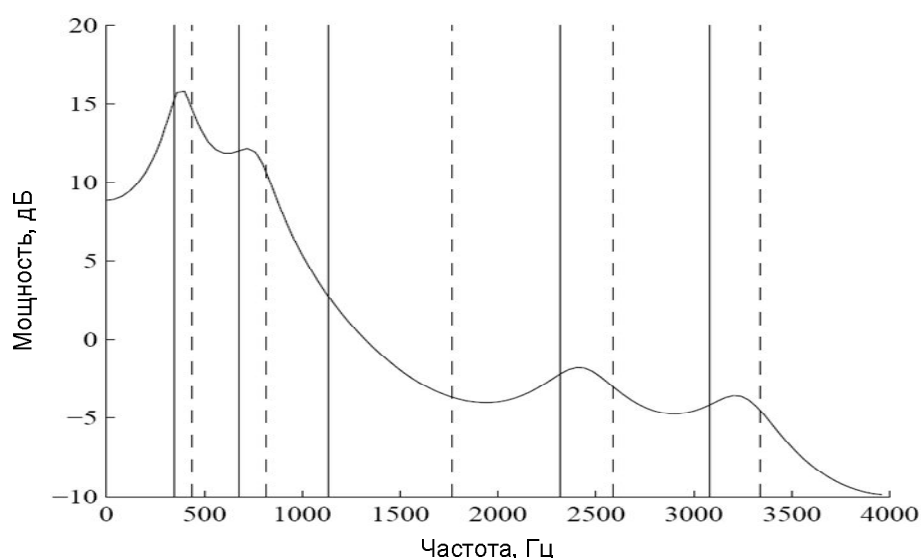


Рис. 8.5 – LPC-спектр мощности с соответствующими LSF (вертикальные линии)

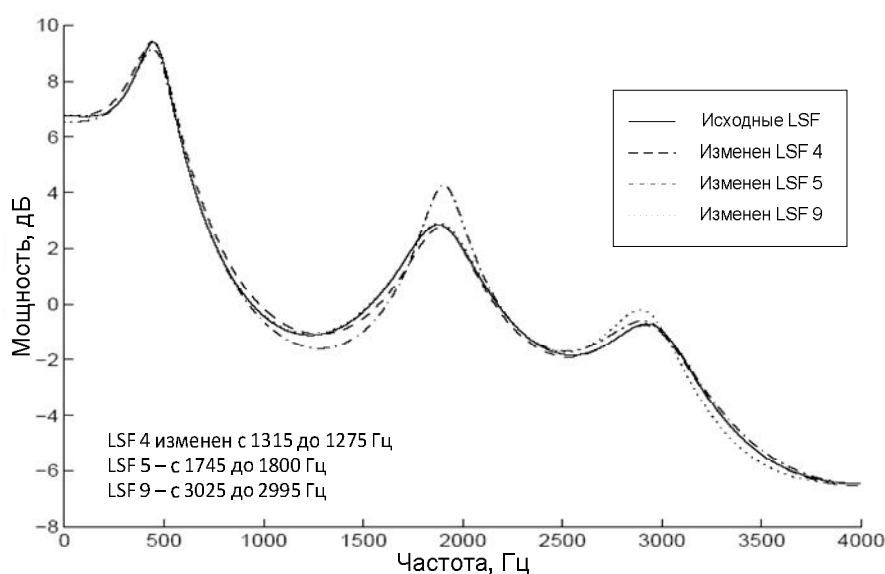


Рис. 8.6 – Влияние локального изменения LSF на LPC-спектр мощности

Наиболее распространенные меры искажений, применяемые при построении квантователя в вокодерах, использующих линейное предсказание, представлены ниже.

1. Логарифмическое искажение спектра. Данная мера искажения используется при оценке качества квантователя каких-либо спектральных характеристик. Рассмотрим спектры мощности $S(\omega)$ и $\hat{S}(\omega)$, соответствующие оригинальному синтезирующему фильтру и модифицированному синтезирующему фильтру (с квантованными коэффициентами). Искажение спектра на фрейме определяется следующим образом [2,3]:

$$sd = \sqrt{1/\pi \int_0^\pi [10 \log_{10} S(\omega) - 10 \log_{10} \hat{S}(\omega)]^2 d\omega}. \quad (8.39)$$

В свою очередь, $S(\omega)$ и $\hat{S}(\omega)$ могут быть определены как

$$S(\omega) = \frac{1}{|A(\omega)|^2}, \quad (8.40)$$

$$\hat{S}(\omega) = \frac{1}{|\hat{A}(\omega)|^2}, \quad (8.41)$$

где

$$|A(\omega)|^2 = \left| 1 + \sum_{k=1}^p a_k e^{-j\omega k} \right|^2 \quad (8.42)$$

является спектром мощности инверсного фильтра. Из вышесказанного следует, что

$$sd = \sqrt{1/\pi \int_0^\pi \left| 10 \log_{10} \frac{|\hat{A}(\omega)|^2}{|A(\omega)|^2} \right|^2 d\omega}. \quad (8.43)$$

2. Взвешенное евклидово расстояние. Данная величина служит математическим критерием для минимизации на стадии квантования, поскольку логарифмическое искажение спектра громоздко для вычисления в реальном времени. Доказано, что взвешенное евклидово расстояние эквивалентно логарифмическому искажению спектра при высокой скорости квантования.

Измерение данной величины можно производить непосредственно в LSF-области, поскольку LSF-коэффициенты хорошо соответствуют спектральной форме. Для выделения отдельной части спектра LSF-коэффициентам, соответствующим этой части, может быть назначен больший вес. Пусть f и \hat{f} – оригинальный и измененный векторы LSF соответственно, тогда их евклидово расстояние $d(f, \hat{f})$ будет определяться следующим образом [1]:

$$d(f, \hat{f}) = \|f - \hat{f}\|^2. \quad (8.44)$$

При использовании LP-анализа порядка p получаем следующее:

$$d(f, \hat{f}) = \sum_{i=1}^p (f_i - \hat{f}_i)^2. \quad (8.45)$$

Взвешенное евклидово расстояние отличается использованием весов:

$$d(f, \hat{f}) = \sum_{i=1}^p [c_i w_i (f_i - \hat{f}_i)]^2, \quad (8.46)$$

где c_i и w_i – веса для i -го LSF-коэффициента. Для фильтра-предсказателя 10-го порядка фиксированные веса c_i определяются следующим образом [3]:

$$c_i = \begin{cases} 1.0, & 1 \leq i \leq 8, \\ 0.8, & i = 9, \\ 0.4, & i = 10. \end{cases} \quad (8.47)$$

Человеческое ухо не способно различить разницу на высоких частотах с такой же точностью, как на низких частотах. Таким образом, эти веса используются для того, чтобы усилить значение низких частот. Адаптивные веса w_i используются для выделения областей спектральной огибающей с большей энергией (формант). Веса w_i определяются как

$$w_i = [S(e^{j\omega})]^r, \quad (8.48)$$

где r – эмпирическая константа.

Для упрощения можно использовать следующую схему вычисления адаптивных весов [2]:

$$w_i = \frac{1}{f_i - f_{i-1}} + \frac{1}{f_{i+1} - f_i}, \quad (8.49)$$

где f_i – LSF-коэффициенты (в радианах), $\omega_0=0$, $\omega_{p+1} = \pi$.

Пример: Дано 30 векторов для тренировки. Необходимо натренировать 4-разрядную кодовую книгу, используя алгоритм K -средних. Ниже на рисунках отображены набор тренировочных векторов и центроиды (обозначены знаком *) на различных этапах тренировки. Границы кластеров показаны условно.

На начальном этапе формируется исходная (начальная) кодовая книга. В данном случае, элементы кодовой книги выбираются из векторов тренировочного множества случайным образом. Затем все элементы тренировочного множества распределяются по кластерам по правилу ближайшего соседа. Как видно из рис. 8.7, тренировочные вектора распределяются по мере близости к центроидам (выбранные ранее элементы кодовой книги). Далее запускается итерационный алгоритм для оптимизации кодовой книги. Для этого в каждом цикле пересчитывается местоположение каждой центроиды и заново перераспределяются все тренировочные вектора. На рис. 8.8 отображены элементы кодовой книги (обозначены знаком *) и множество тренировочных векторов после первой итерации. Как видно из рис. 8.8, по сравнению с предыдущим состоянием изменились не только местоположения центроид, но и изменилось количество элементов в кластерах (произошло перераспределение – в каком-то кластере количество элементов увеличилось, а в каком-то уменьшилось). Таким образом, в процессе трени-

ровки на каждой итерации пересчитываются местоположения центроид и перераспределяются элементы от кластера к кластеру (рис. 8.9) до тех пор, пока будет происходить уменьшение общего искажения. Когда вновь полученная ошибка квантования практически не будет отличаться от предыдущей, процесс оптимизации прерывается.

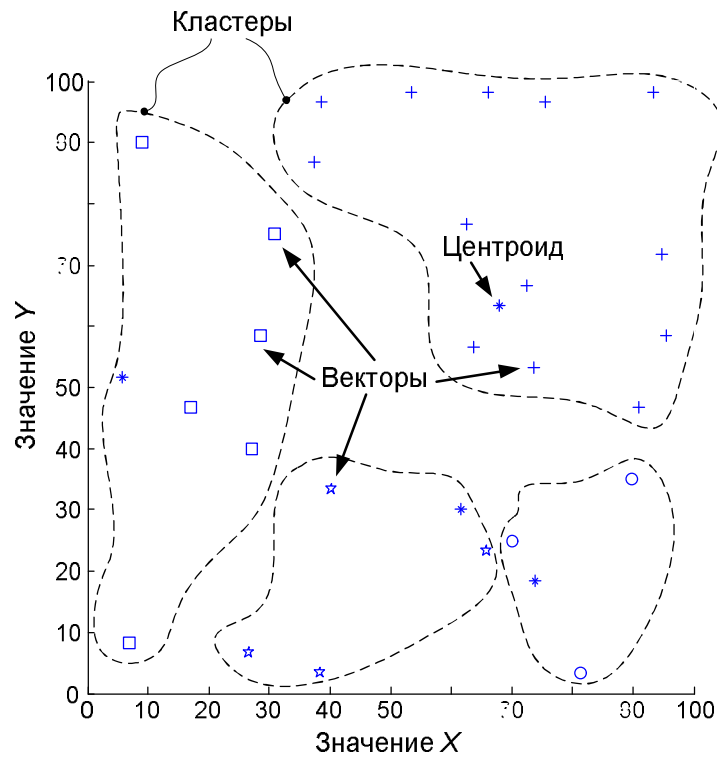


Рис. 8.7 – Формирование исходной кодовой книги

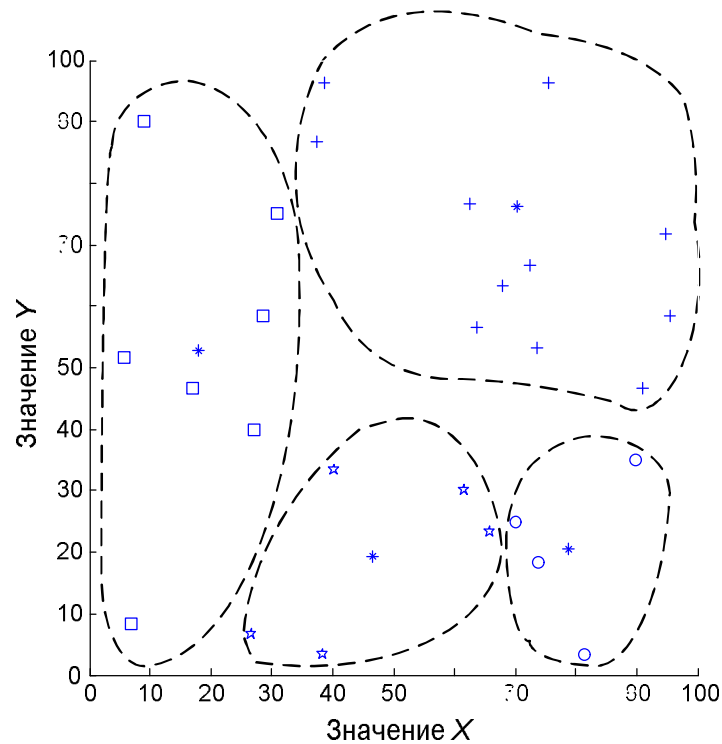


Рис. 8.8 – Расположение кластеров после первой итерации

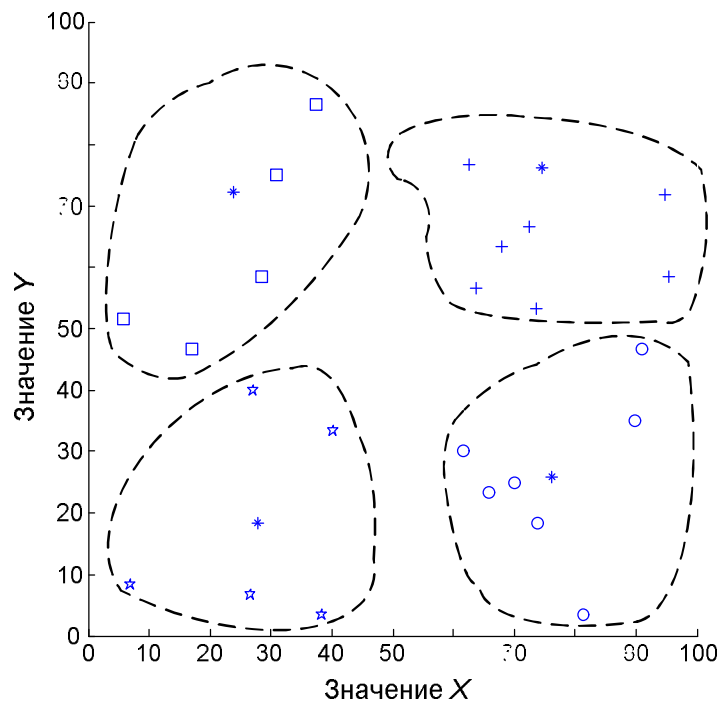


Рис. 8.9 – Оптимальная кодовая книга

8.7 Квантование амплитуд гармоник речевого сигнала с использованием закономерностей психоакустики

Одним из фундаментальных вопросов в приложениях кодирования на базе гармонических моделей является квантование амплитуд гармоник, так как качество реконструированной речи в параметрических вокодерах в большой степени зависит от качества квантования параметров гармонической компоненты, несущей основную информацию о кодируемом речевом сигнале.

В настоящее время известно достаточно большое количество подходов кодирования последовательности амплитуд гармоник. Скалярное квантование, например, квантует каждый элемент индивидуально; тем не менее, векторное квантование [1] является более предпочтительным подходом для современных алгоритмов низкоскоростных кодеров речи, что обусловлено лучшим соотношением качество / скорость передачи. Традиционные векторные квантователи строятся с учетом фиксированной длины векторов. В последних работах удалось добиться достаточно высокого качества квантования гармонических амплитуд благодаря применению схемы расщепленного векторного квантования линейных спектральных пар, при этом прозрачное кодирование достигалось при скорости 23 бит/вектор [5]. Тем не менее, построение векторного квантователя с переменной длиной кодируемого вектора амплитуд гармоник выглядит более естественным решением ввиду того, что при этом не требуется дополнительных преобразований над входным вектором.

В контексте гармонической модели проблема квантования в большей степени связана с передачей вектора амплитуд гармоник. Если рассмотреть изменение спектра речевого сигнала во времени для разных дикторов (рис. 8.10), можно сделать вывод, что векторы амплитуд

гармоник, даже определяющие голос одного и того же диктора, имеют различную размерность в разные моменты времени.

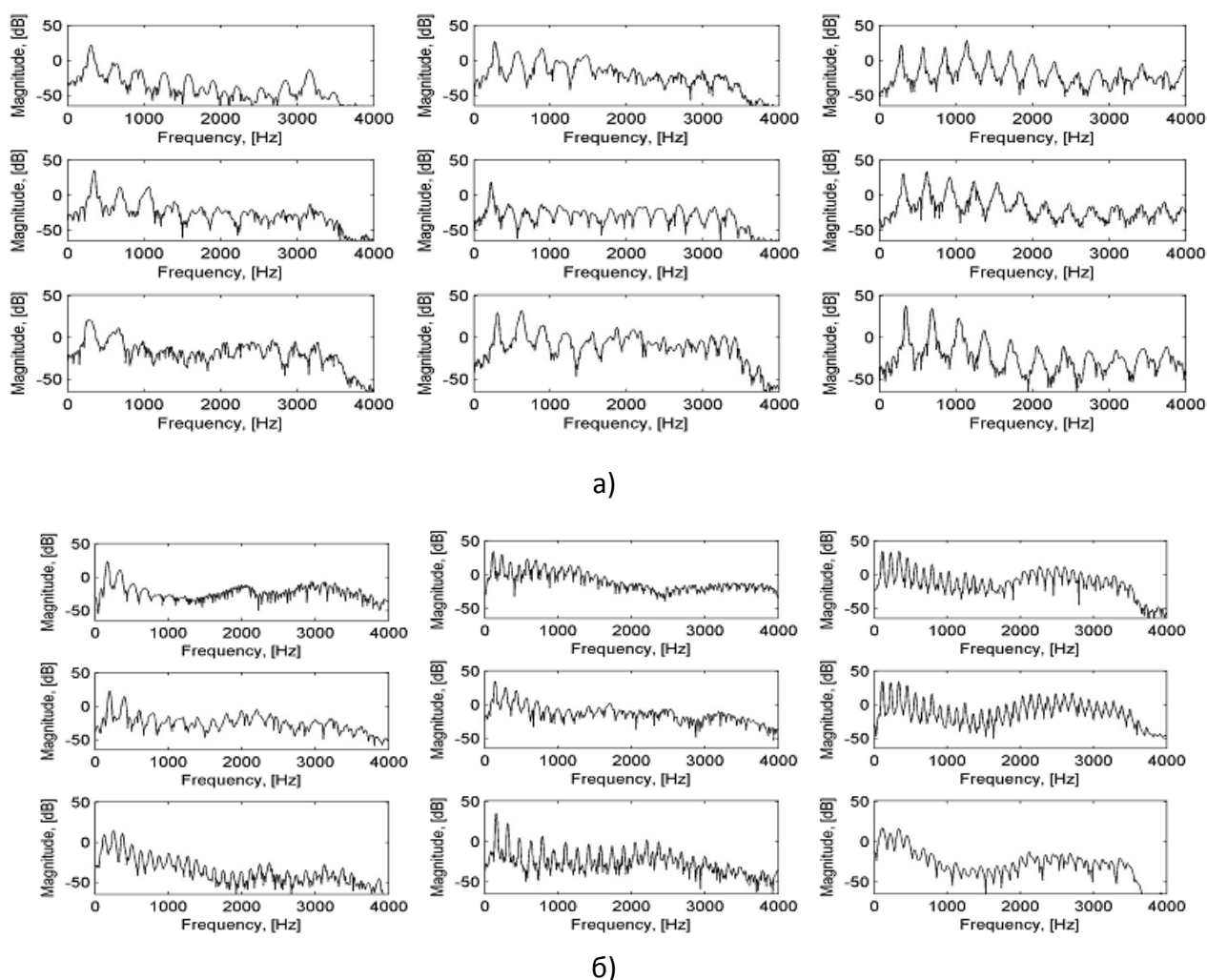


Рис. 8.10 – Изменение спектра речи во времени: а) женский голос; б) мужской голос

К сожалению, математический аппарат векторного квантования был разработан для квантования векторов фиксированной длины и практически не используется с векторами переменной длины, такими как векторы амплитуд гармоник. Для решения данной проблемы возможны различные подходы. Одним из вариантов является использование собственной кодовой книги для каждой размерности [6]. Естественно, такой подход является малоприемлемым для использования в системах реального времени из-за серьезных требований к объему памяти. В наиболее широко применяемых решениях осуществляются различные преобразования над векторами переменной размерности с тем, чтобы привести их размерность к некоторому заданному фиксированному значению (с сохранением формы речевого спектра) с последующим применением техник векторного квантования. Примерами таких решений могут служить [5,7–9]. Очевидным недостатком здесь является необходимость дополнительных преобразований и, следовательно, возможность внесения дополнительных искажений.

Одно из возможных решений – квантование фиксированного количества гармонических амплитуд, например в кодере на базе линейного предсказания со смешанным возбуждением (MELP – Mixed Excitation Linear Prediction) [10] векторное квантование используется для

квантования первых 10 гармонических амплитуд, а амплитуды остальных гармоник считаются равными амплитуде последней (10-й) гармоники. Легко заметить, что 10 гармоник покрывают весь или почти весь речевой спектр для женских голосов с высокой частотой основного тона, в то время как для мужских голосов они могут покрыть только одну четвертую всего частотного диапазона (рис. 8.10, а, б), что означает существенную потерю качества для мужских голосов по сравнению с женскими.

Наконец, в [11] была разработана схема векторного квантования с переменной размерностью векторов (от англ. – Variable Dimension Vector Quantization – VDVQ). Тем не менее, поскольку в этом подходе не учитываются закономерности психоакустики, его трудно считать оптимальным.

Далее будет рассмотрен математический аппарат VDVQ и его модификация с точки зрения восприятия речи человеком.

8.7.1 Векторное квантование с переменной размерностью векторов

В схеме VDVQ, предложенной в [11], кодовая книга квантователя содержит N_c кодовых векторов y_i , $i = 0, \dots, N_c - 1$, при

$$y_i^T = [y_{i,0} \ y_{i,1} \ \wedge \ y_{i,N_V-1}], \quad (8.50)$$

где N_V – размерность кодового вектора.

Пусть поиск вектора гармонических амплитуд x с размерностью $N(\omega_0)$ и нормализованной частотой основного тона ω_0 осуществляется путем полного перебора в кодовой книге, тогда требуется рассчитать расстояния $d_i(x, \hat{y}_i)$, $i = 0, \dots, N_c - 1$, где

$$\hat{y}_i^T = [\hat{y}_{i,1} \ \hat{y}_{i,2} \ \wedge \ \hat{y}_{i,N(\omega_0)}], \quad \hat{y}_{i,j} = y_{i,k_j}, \quad j = 1, \dots, N(\omega_0), \quad (8.51)$$

при

$$k_j = \left[N_V \omega_j / \pi \right], \quad \omega_j = j \omega_0, \quad j = 1, \dots, N(\omega_0), \quad (8.52)$$

где $[\]$ в данном случае означает округление к ближайшему целому.

Схема работает следующим образом: для каждого кодового вектора y_i путём расчета набора индексов k_j извлекается вектор \hat{y}_i , имеющий ту же размерность, что и x . Эти индексы рассчитываются в соответствии с периодом основного тона и указывают на элементы y_i , ближайшие к позиции j -й гармоники в кодовой книге. После расчета всех расстояний d_i для квантования x выбирается индекс кодового вектора с наименьшим расстоянием. В качестве расстояния (меры искажения) используется спектральное отклонение:

$$SD = \sqrt{\frac{1}{N(\omega_0)} \sum_{j=1}^{N(\omega_0)} (x_j - \hat{y}_j)^2}. \quad (8.53)$$

Улучшенная конфигурация схемы VDVQ, называемая IVDVQ, предложена в [12]. Улучшение основано на интерполяции элементов кодовых векторов y_i для получения действительных кодовых векторов \hat{y}_i . Индексы k_j в IVDVQ рассчитываются без операции округления:

$$k_j = \frac{N_V \omega_j}{\pi}, \quad \omega_j = j \omega_0, \quad j = 1, \dots, N(\omega_0). \quad (8.54)$$

Элемент $\hat{y}_{i,j}$ получается путем линейной интерполяции между двумя элементами вектора y_i , определяемыми индексами $[k_j]$ и $[k_{j+1}]$:

$$\hat{y}_{i,j} = y_{i,[k_j]} + \{k_j\} (y_{i,[k_{j+1}]} - y_{i,[k_j]}), \quad (8.55)$$

где $\{k_j\}$ обозначает дробную часть выражения (8.54). Обучение кодовых книг по методам VDVQ и IVDVQ представляет собой вариацию на тему алгоритма “ k -средних” [8] и подробно описано в [12]. Результат применения метода к квантованию амплитуд гармоник отражён на рис. 8.11, использовалась 10-разрядная кодовая книга.

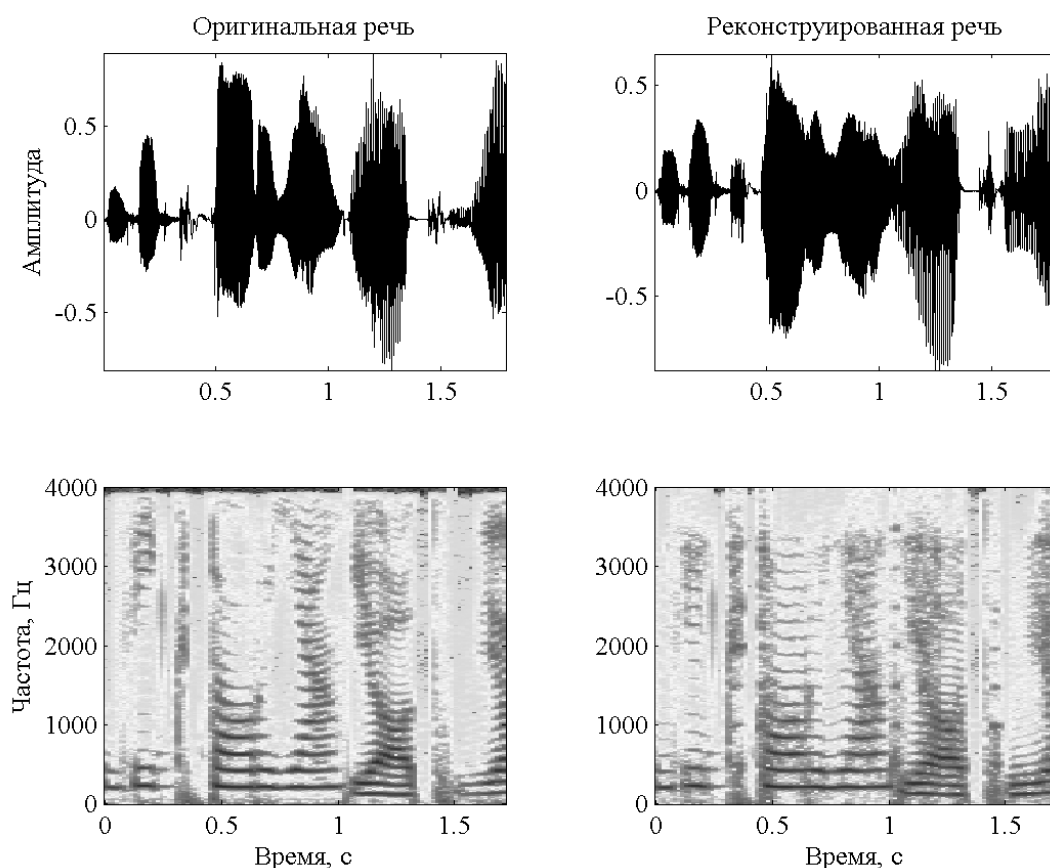


Рис. 8.11 – Пример восстановления речи, кодированной с использованием метода VDVQ

8.7.2 VDVQ с психоакустически обоснованным ограничением длины вектора

Кодовые книги для VDVQ-метода обычно имеют большую длину кодовых слов (от 41 до 109 – в экспериментах [12]), что приводит к высоким требованиям к объёму памяти для их хранения. В то же время, можно видеть, что последние гармонические амплитуды спектра имеют незначительную величину, особенно в случае мужской речи (рис. 8.10б). Следовательно, имеет смысл ограничить размерность квантуемого вектора таким образом, чтобы не учитывать

достаточно малые амплитуды. Схожая проблема существует в рамках модели речевого сигнала “гармоники плюс шум” [13,14], где необходимо найти максимальную частоту вокализованности (ограничить спектр гармонической компоненты). Алгоритм, предложенный в [14], осуществляет проверку спектра на “гармоничность” в окрестности гармонических амплитуд, в случае, если спектр в области двух смежных проверяемых гармоник оказался “негармоническим”, проверка прекращается. В качестве максимальной частоты вокализованности принимается последняя гармоника частоты основного тона, предшествующая “негармонической” области спектра. Всё же данный алгоритм является в большой степени эвристическим и использует при оценке некоторые заранее определённые опытным путём пороговые значения.

Используемая модель анализа речевого сигнала [15] предполагает разделение речи на гармоническую и шумовую компоненту по всему спектру. Используя закономерности психоакустики, можно определить, в какой степени шумовая компонента влияет на восприятие человеком гармонической компоненты, т.е. определить гармоники, не влияющие на восприятие речи в целом [15,16].

Для решения данной проблемы использовалась психоакустическая модель Джонсона [17]. Данная модель позволяет рассчитать порог маскирования “шум маскирует тон” в частотной области с использованием следующей последовательности действий:

- 1) сегмент шумовой компоненты [15] взвешивается временным окном и подвергается ДПФ;
- 2) спектр мощности шумовой компоненты суммируется в критических полосах, измеряемых в барках [18]:

$$B_i = \sum_{n=bl_i}^{bh_i} P(n), \quad (8.56)$$

где $P(n)$ – n -й частотный компонент спектра мощности; bl_i, bh_i – номер начального и конечного спектрального отсчёта, попадающих в i -ю критическую полосу.

Шкала барков получается с помощью следующего преобразования:

$$z(f) = 1 + 13 \operatorname{arctg}(0,76f) + 3,5 \operatorname{arctg} \left[\left(\frac{f}{7,5} \right)^2 \right], \quad (8.57)$$

где f – частота в Гц. Для ДПФ размерности 256 и частоты дискретизации $F_s = 8000$ Гц параметры критических полос приведены в табл. 8.1.

- 3) рассчитывается функция распространения для оценки эффектов маскирования в нескольких критических полосах [19]:

$$S_{i,j} = 10^{\frac{(15,81+7,5(k+0,474))-17,5\sqrt{1+(k+0,474)^2}}{10}}. \quad (8.58)$$

где $k = i - j$, i – номер барка маскируемого сигнала; j – номер барка маскирующего сигнала.

- 4) вычисляется распространение спектральной энергии барка в каждой критической полосе как свёртка B_i с функцией распространения $S_{i,j}$:

$$\begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_{18} \end{bmatrix} = \begin{bmatrix} S_{1,1} & S_{1,2} & S_{1,3} & \dots & S_{1,18} \\ S_{2,1} & S_{2,2} & S_{2,3} & \dots & S_{2,18} \\ S_{3,1} & S_{3,2} & S_{3,3} & \dots & S_{3,18} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ S_{18,1} & S_{18,2} & S_{18,3} & \dots & S_{18,18} \end{bmatrix} \times \begin{bmatrix} B_1 \\ B_2 \\ B_3 \\ \vdots \\ B_{18} \end{bmatrix}. \quad (8.59)$$

5) рассчитываются коэффициенты тональности для каждой критической полосы:

$$\alpha_i = \min\left(\frac{SFM_{dB}(i)}{SFM_{dBmax}}, 1\right), \quad (8.60)$$

где $SFM_{dB}(i)$ – мера спектральной пологости в i -ой критической полосе:

$$SFM_{dB} = 10[\log_{10}(GM) - \log_{10}(AM)], \quad (8.61)$$

где AM и GM – среднее арифметическое и среднее геометрическое значение спектра мощности в i -ой критической полосе; SFM_{dBmax} – максимальное значение меры спектральной пологости равно -60 дБ.

Табл. 8.1 – Параметры критических полос приведены для ДПФ размерности 256 и частоте дискретизации $F_s = 8000$ Гц

Номер критической полосы	Номера элементов ДПФ	Количество элементов ДПФ	Частоты, Гц
1	1...3	3	0...94
2	4...6	3	94...187
3	7...10	4	187...312
4	11...13	3	312...406
5	14...16	3	406...500
6	17...20	4	500...625
7	21...25	5	625...781
8	26...29	4	781...906
9	30...35	6	906...1094
10	36...41	6	1094...1281
11	42...47	6	1281...1469
12	48...55	8	1469...1719
13	56...64	9	1719...2000
14	65...74	10	2000...2312
15	75...86	12	2312...2687
16	87...100	14	2687...3125
17	101...118	18	3125...3687
18	119...128	9	3687...4000

6) определяются смещения порогов маскирования:

$$O_i = 5,5(1 - \alpha_i). \quad (8.62)$$

7) производится расчёт порогов маскирования в критических полосах и их ренормализация:

$$T_i = 10^{\log_{10}(C_i) - O_i/10}. \quad (8.63)$$

Для ренормализации требуется определить ошибку распространения спектральной энергии барка, для этого предполагается, что на слуховую систему воздействует гипотетический раздражитель, спектральная энергия которого в критической полосе равна единице:

$$\begin{bmatrix} C_{E1} \\ C_{E2} \\ C_{E3} \\ \vdots \\ C_{E18} \end{bmatrix} = \begin{bmatrix} S_{1,1} & S_{1,2} & S_{1,3} & \dots & S_{1,18} \\ S_{2,1} & S_{2,2} & S_{2,3} & \dots & S_{2,18} \\ S_{3,1} & S_{3,2} & S_{3,3} & \dots & S_{3,18} \\ \vdots & \vdots & \vdots & & \vdots \\ S_{18,1} & S_{18,2} & S_{18,3} & \dots & S_{18,18} \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}. \quad (8.64)$$

Ренормализованные пороги маскирования определяются как

$$T'_i = T_i - 10 \log_{10}(C_{Ei}). \quad (8.65)$$

8) окончательные значения порогов маскирования определяются как

$$T_i^f = \max(T'_i, ATH(f)). \quad (8.66)$$

где $ATH(f)$ – функция, аппроксимирующая значение абсолютного порога слышимости [18], рассчитывается с помощью выражения (8.67) для частот, равных значениям гармоник частоты основного тона:

$$ATH(f) = 3,64f^{-0,8} - 6,5e^{-0,6(f-3,3)^2} + 10^{-3}f^4. \quad (8.67)$$

где f – частота в кГц.

Максимальной частотой вокализованности считается последняя гармоника частоты основного тона, превышающая порог маскирования.

На рис. 8.13 показан результат расчёта порога маскирования и определения максимальной частоты вокализованности для вектора гармонических амплитуд. Очевидно, что вычислительная сложность поиска в кодовой книге в данном случае будет снижена более чем в 2 раза.

Таким образом, удаётся ограничить размерность вектора гармонических амплитуд на основании закономерностей психоакустики и тем самым снизить вычислительную сложность процесса квантования гармонических амплитуд. Результат применения метода отражён на рис. 8.13, использовалась 10-разрядная кодовая книга.

Метод квантования векторов переменной размерности является весьма удобным для использования с такими параметрами гармонической модели речи как амплитуды, поскольку отпадает надобность в дополнительных преобразованиях. Предложенные методы, в основе которых лежат преобразования, использующие закономерности психоакустики позволяют повысить качество реконструированной речи и снизить вычислительную сложность алгоритмов квантования.

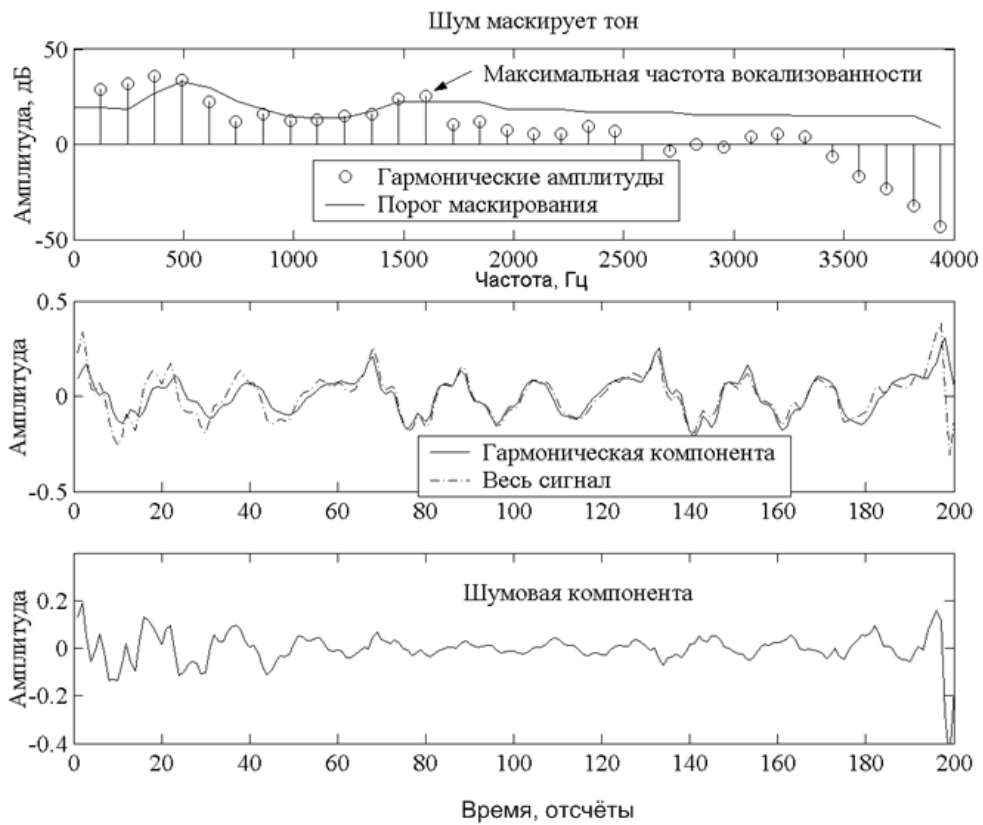


Рис. 8.12 – Иллюстрация маскирования амплитуд гармоник

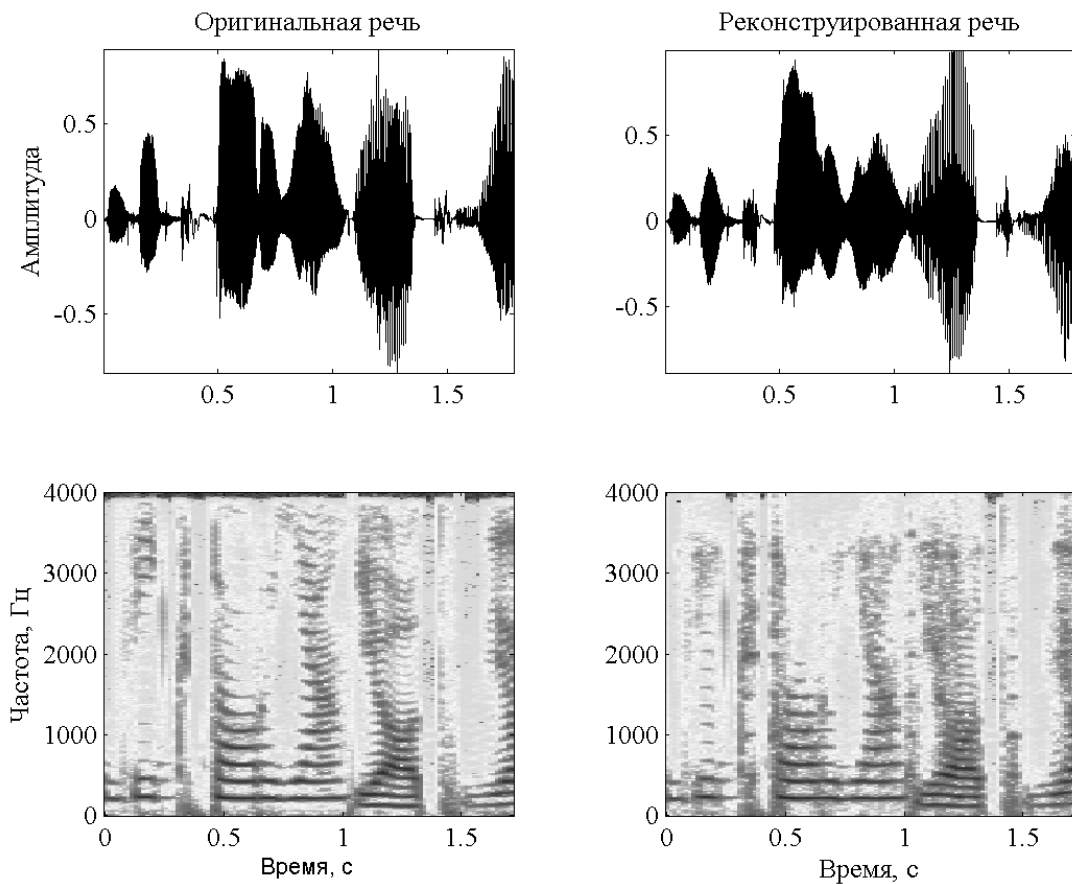


Рис. 8.13 – Результат применения метода VDVQ с психоакустически мотивированным ограничением длины вектора

8.8 Список использованных источников

1. Gersho A., Gray R.M., Vector Quantization and Signal Compression, Boston: Kluwer Academic Press, 1992.
2. Kleijn W.B., Palival K.K., eds. Speech coding and synthesis, Amsterdam: Elsevier, 1995.
3. Kondoz A.M. Digital speech: coding for low bit rate communication systems, NY.: John Wiley & Sons, Inc., 1996, 442 p.
4. Рабинер Л.Р., Шафер Р.В., Цифровая обработка речевых сигналов, М.: Радио и связь, 1981, 495 с.
5. Павловец А., Петровский А., Квантование огибающей спектра в вокоде, основанное на декомпозиции речевого сигнала на периодическую и аperiodическую составляющие, Цифровая обработка сигналов, № 3, 2005 г., с. 13-21.
6. Adoul J.P., Delprat M., Design algorithm for variable-length vector quantizers, Proc. Allerton Conf. on Circuits, Syst., Comput., 1986, pp. 1004–1011.
7. Li C., Lupini P., Shlomot E., and Cuperman V., Coding of variable dimension speech spectral vectors using weighted nonsquare transform vector quantization, IEEE Trans. Speech, and Audio Processing, vol. 9, no. 6, 2001, pp. 622-631.
8. MacQueen J.B., Some Methods for Classification and Analysis of Multivariate Observations, Proc. Fifth Berkeley Symp. Math. Statistics and Probability, vol. 1, 1967, pp. 281-296.
9. McAulay R.J., Quatieri T.F., Sinusoidal Coding in Speech Coding and Synthesis, (W.Klein and K. Palival, eds.), Amsterdam: Elsevier Science Publishers, 1995, pp. 121–176.
10. Supplee L., Cohn R., Collura J., McCree A., MELP: the new federal standard at 2400 bps, Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'97, vol. 2, Munich, Germany, April 1997, pp. 1591–1594.
11. Das A., Rao A., Gersho A., Variable-dimension vector quantization, IEEE Signal Processing Letters, vol. 3, no. 7, 1996, pp. 200-202.
12. Chu W.C., A novel approach to variable dimension vector quantization of harmonic magnitudes, Proc 3rd IEEE International Symposium on Image and Signal Processing and Analysis, vol. 1, Rome, Italy, September 2003, pp. 537-542.
13. Bao C., Lukasiak J., Ritz C., A novel voicing cut-off determination for low bit-rate harmonic speech coding, in INTERSPEECH-2005, pp. 2709-2712.
14. Stylianou Y., Applying the harmonic plus noise model in concatenative speech synthesis”, IEEE Transactions on Speech and Audio Processing, vol. 9, no.1, Jan. 2001, pp. 21 – 29.
15. Павловец А., Зубрыцки П., Петровский А., Гармоническая модель речевого сигнала: определение параметров и их квантование, Доклады БГУИР, 2007, № 4 (20), с. 19 – 34.

-
16. Павловец А., Петровский А., Использование закономерностей психоакустики в процедуре квантования параметров гармонической модели речевого сигнала, Речевые технологии, 2008, №4, с. 50 – 60.
 17. Johnston J., Estimation of perceptual entropy using noise masking criteria, Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'88, vol. 5, New York, NY, USA, April 1988, pp. 2524-2527.
 18. Zwicker E., Fastl H., Psychoacoustics: facts and models, Springer-Verlag, Berlin. 1990.
 19. Schroeder M. R., Atal B. S., Hall J. L., Optimizing digital speech coders by exploiting masking properties of the human ear, Journal of the Acoustical Society of America, vol. 66, pp. 1647-1652, 1979.

9 ПРИМЕНЕНИЕ ГИБРИДНОГО ПАРАМЕТРИЧЕСКОГО ОПИСАНИЯ ДЛЯ КОДИРОВАНИЯ ЗВУКОВЫХ И РЕЧЕВЫХ СИГНАЛОВ

Петровский Ал.А., Азаров И.С., Петровский А.А.

АННОТАЦИЯ

В данном разделе рассматривается гибридное описание сигналов, основанное на комбинированном использовании нескольких моделей параметрического представления. Рассматриваются вопросы, связанные с разделением сигнала на различные составляющие для достижения максимальной эффективности подхода. Описываются все необходимые математические и алгоритмические средства, и приводится результат их реализации.

Ключевые слова – гибридное описание сигнала, мгновенные гармонические параметры, периодический сигнал, шумовой сигнал, переходной сигнал.

СОДЕРЖАНИЕ

9.1	Введение.....	190
9.2	Общая структура системы гибридного описания	191
9.3	Синусоидальное и гармоническое представления сигнала.....	192
9.4	Оценка мгновенных гармонических параметров	194
9.4.1	Гармонический анализ звукового сигнала	194
9.4.2	Гармонический анализ речевого сигнала.....	198
9.5	Моделирование переходных составляющих сигнала методом согласованной подгонки	202
9.5.1	Согласованная подгонка с динамически оптимизируемым словарем на основе психоакустики	202
9.5.2	Определение переходных фреймов..	205
9.6	Применение гибридной системы анализа.....	205
9.7	Заключение.....	207
9.8	Список использованных источников.....	207

9.1 Введение

Цель данного раздела – изложить новые методы для гибридного описания звуковых и речевых сигналов, которые позволяют обеспечить более высокое качество реконструкции за счет более точного разделения сигнала на составляющие различной природы. Современные системы для обработки и кодирования звука и речи различают три основных компонента: пе-

риодическую, шумовую и переходную (транзиентную) составляющие. Первая из них представляет собой сумму тригонометрических функций (синус и косинус); вторая – нерегулярный сигнал без четкой локализации спектра, который на слух воспринимается как шум; третья – непериодические (часто непродолжительные) фрагменты с резким изменением энергии и формы. Гибридные модели позволяют повысить эффективность кодирования сигналов за счет использования психоакустических критериев оценки, а так же применения специальных способов шумоподавления и поэтому широко используются в современных системах компрессии [1-9]. Аналогичное гибридное параметрическое представление используется в системах обработки звука для создания специальных эффектов (изменение масштаба времени, изменение основного тона, тембра и др.) [10,11], и в системах синтеза речи по тексту [12,13].

Представление речевого сигнала при помощи гибридной модели представляет особый интерес, так как фрагменты речи различной природы звукообразования (вокализованные и невокализованные) относятся к отдельным составляющим модели (периодической и остаточной соответственно). При гармоническом анализе речи сигнал разделяют на вокализованные и невокализованные фрагменты, применяя синусоидальное описание только к первым из них. Вокализованная речь часто моделируется при помощи синусоид, гармонически расположенных в частотной области (гармоническая модель), что значительно сокращает число параметров, необходимых для хранения, обработки или передачи.

Модель гибридного описания была введена в работе [14] с целью повышения качества широкополосных кодеров с низким битрейтом. Модель является очень удачной, поскольку позволяет эффективным образом описать практически любой слышимый сигнал, однако ее потенциал продолжает раскрываться с совершенствованием методов анализа и сепарации сигналов. Особого внимания заслуживают новые методы мгновенного синусоидально-гармонического анализа для выделения и обработки периодической части сигнала и новые методы согласованной подгонки на основе вейвлет пакетов для описания переходной составляющей, поскольку их совместное применение в системе гибридного описания может существенно повысить качество реконструированного сигнала в системах кодирования и обработки звуковой информации [15].

9.2 Общая структура системы гибридного описания

Как было сказано выше, подход гибридного моделирования звуковых и речевых сигналов основан на комбинации трех различных средств параметрического представления, используемых в цифровой обработке сигналов: синусоидальное моделирование, согласованная подгонка с психоакустически оптимизированным для каждого фрейма словарем вейвлет пакета и адаптированное к шкале барков моделирование шума. Синусоидальная часть представляется в виде суммы синусоид с мгновенными параметрами (амплитуда, частота и фаза), переходная составляющая моделируется при помощи согласованной подгонки, а шум обрабатывается на основе барк-адаптированных вейвлет пакетов. Общая структура системы гибридного анализа показана на рис. 9.1.

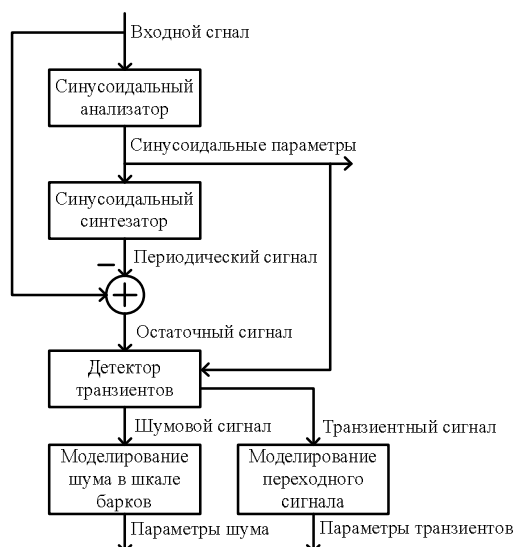


Рис. 9.1 – Общая структура система гибридного анализа

К входному сигналу применяется синусоидальное моделирование, затем в остаточном сигнале выполняется определение транзиентов (переходных фреймов) путем энергетического порога в сочетании с методом выделения границы относительной громкости. После того, как синусоидальные и переходные компоненты выделены из сигнала, оставшаяся (шумовая) часть сигнала описывается при помощи модели на основе барк адаптивных вейвлет пакетов. Транзиентная же часть описывается при помощи согласованной подгонки.

Вышеприведенная схема анализа обеспечивает хорошую координацию используемых моделей и позволяет эффективно обрабатывать любой речевой или звуковой сигнал.

9.3 Синусоидальное и гармоническое представления сигнала

Синусоидальное (гармоническое) представление сигнала широко применяется в современных системах обработки звука, так как с его помощью можно эффективно описывать периодические компоненты, которые составляют значительную часть различных звуков и речи. Синусоидальная и гармоническая модели были с успехом использованы в системах кодирования речи, таких как мультиполосный вокодер [16-18] и трансформ кодер [19], а так же в кодировании широкополосных звуковых сигналов [11,20,21].

Хотя предположение о том, что произвольная функция может быть представлена в виде суммы синусоидальных составляющих принадлежит Жану Батисту Фурье и была высказана еще в 1807 году, модель для представления дискретных звуковых сигналов в виде суммы синусоидальных компонент с изменяющимися во времени амплитудами, частотами и фазами была представлена сравнительно недавно [22]. В работе [23] представлена модель, которая рассматривает вокализованную речь как сумму гармонически связанных синусоид с амплитудами и фазами, вычисляемыми непосредственно из спектра дискретного преобразования Фурье, а невокализованная речь рассматривается как сумма случайно распределенных синусоид со случайными начальными фазами. Позже данная модель была значительно усовершенствована [10].

Синусоидальное представление предполагает, что моделируемый сигнал является комбинацией синусоид с медленно изменяющимися амплитудами и частотами и, следовательно, не всегда эффективно для описания шумовых и взрывных звуков. Таким образом, одна из особенностей использования синусоидальных представлений применительно к звуковым сигналам заключается в том, что часто нет необходимости точно описывать весь исходный сигнал в виде суммы синусоид. Намного эффективнее с практической точки зрения разделить сигнал на отдельные составляющие, для каждой из которых применить затем свою модель описания [11,24-26]. Так как синусоидальное моделирование в большинстве случаев можно применить непосредственно к входу системы, входной сигнал удобно рассматривать как комбинацию периодической и остаточной компонент [27]. Таким образом, сигнал $s(n)$ можно записать в виде соотношения:

$$s(n) = \sum_{k=1}^K \text{MAG}_k(n) \cos \varphi_k(n) + r(n), \quad (9.1)$$

где $\text{MAG}_k(n)$ – мгновенная амплитуда k -й синусоиды; K – число синусоид; $\varphi_k(n)$ – мгновенная фаза k -й синусоиды; $r(n)$ – сигнал-остаток.

Мгновенная фаза $\varphi_k(n)$ и мгновенная частота $f_k(n)$ соотносятся следующим образом:

$$\varphi_k(n) = \sum_{i=0}^n \frac{2\pi f_k(i)}{F_s} + \varphi_k(0), \quad (9.2)$$

где F_s – частота дискретизации; $\varphi_k(0)$ – начальная фаза k -й синусоиды.

Наряду с синусоидальной моделью широко применяется (особенно в вокодерных системах) гармоническая модель, которая предполагает, что значения мгновенных частот $f_k(n)$ являются кратными частоте основного тона $f_0(n)$ и могут быть вычислены по следующей формуле:

$$f_k(n) = k f_0(n). \quad (9.3)$$

Гармоническая модель используется для кодирования речевых сигналов с высоким коэффициентом сжатия [14,28-31], так как обеспечивает чрезвычайно эффективное описание вокализованных фрагментов речи.

Задачей гармонического анализа является определение синусоидальных параметров $\text{MAG}_k(n)$, $f_k(n)$ и $\varphi_k(n)$ для заданного момента или интервала времени. Считается, что амплитуда и фаза синусоидальных компонент изменяется медленно, поэтому можно сделать следующие предположения:

- каждая синусоида может быть ограничена в частотной области узкой частотной полосой;
- синусоидальные компоненты разделены в частотной области (их можно выделить на всем протяжении анализируемого фрейма фильтрами с неперекрывающимися полосами пропускания), в противном случае они создаются переходной (транзиентной) компонентой;
- синусоидальные компоненты достаточно продолжительны, в противном случае они формируют либо переходной компонент, либо шум.

Таким образом, искомые параметры синусоидальной модели $MAG_k(n)$ и $f_k(n)$ являются гладкими, непрерывными функциями с ограниченным частотным диапазоном.

Разделение сигнала на периодическую и остаточную части, как и оценка гармонических параметров, является фундаментальной задачей синусоидального моделирования. Точность оценок, как правило, оказывает существенное влияние на качество работы систем, что свидетельствует о необходимости совершенствования методов гармонического анализа. Неточное разделение вносит в обрабатываемый сигнал слышимые артефакты, которые затем, на последующих стадиях обработки, не могут быть исключены. В качестве основного инструмента для выполнения гармонического анализа используется дискретное преобразование Фурье (ДПФ), что накладывает определенные ограничения. В этом случае предполагается, что анализируемый сигнал является квазистационарным, то есть на протяжении некоторого периода времени его параметры остаются неизменными. Несмотря на то, что, используя ДПФ, были получены достаточно неплохие результаты в системах синтеза речи по тексту [12] и кодирования речи [32], все же допущение локальной стационарности сигнала значительно сдерживает дальнейшее развитие данных систем. Прежде всего, используя ДПФ, сложно получить адекватное параметрическое описание неустойчивых тональных звуков. Другая проблема заключается в сложности анализа сигналов с быстро изменяющимся тоном. Например, ДПФ очень ограничено применимо к оценке параметров гармоник высокого порядка вокализованной речи из-за свойственного им быстрого изменения частоты.

9.4 Оценка мгновенных гармонических параметров

9.4.1 Гармонический анализ звукового сигнала

Схема гармонического анализатора звуковых сигналов для оценки мгновенных параметров, выполненного на основе узкополосной фильтрации, представлена на рис. 9.2.

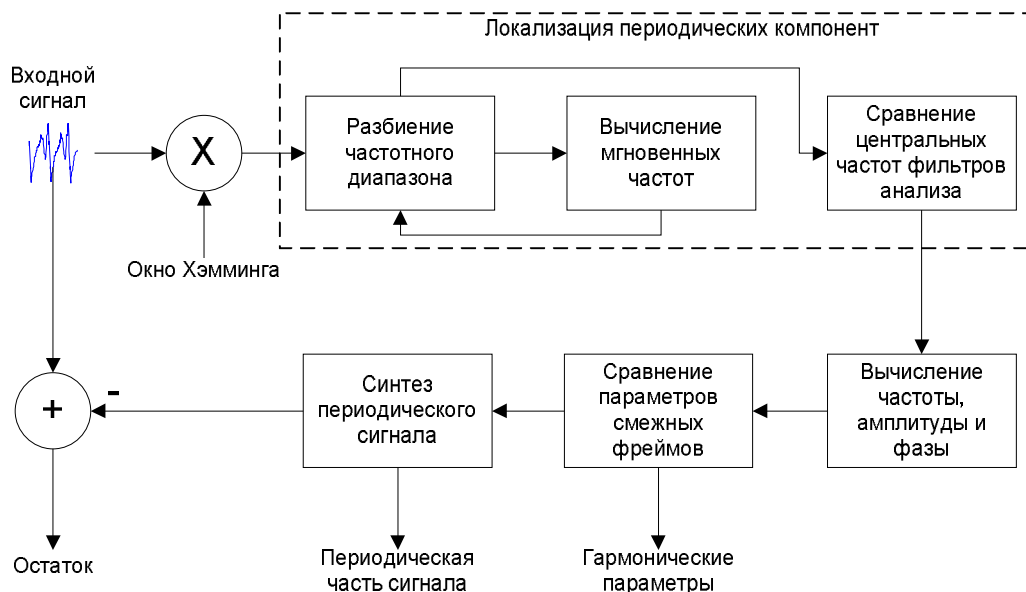


Рис. 9.2 – Схема гармонического анализатора звукового сигнала

Входной сигнал анализируется и разделяется на периодическую компоненту и остаток, которые вместе с полученными гармоническими параметрами передаются последующим блокам анализа и кодирования. Блоки могут быть различными в зависимости от схемы звукового кодера.

В соответствии с применяемой схемой анализа входной сигнал разбивается на фреймы с перекрытиями, которые анализируются отдельно друг от друга при помощи узкополосной фильтрации. Полученные гармонические параметры смежных фреймов сравниваются при помощи буфера слежения для выявления длинных и стабильных гармонических компонент. Используя полученные параметры, синтезируется синусоидальный сигнал, который вычитается из исходного для получения остатка.

В работе [15] был предложен специальный частотно-модулированный (ЧМ) фильтр анализа, для оценки мгновенных гармонических параметров. Оценка параметров сводится к вычислению мгновенной амплитуды $MAG(n)$, фазы $\varphi(n)$ и частоты $f(n)$, при помощи следующих выражений:

$$MAG(n) = \sqrt{A^2(n) + B^2(n)}; \quad (9.4)$$

$$\varphi(n) = \arctan\left(\frac{-B(n)}{A(n)}\right); \quad (9.5)$$

$$f(n) = \frac{\varphi(n+1) - \varphi(n)}{2\pi} F_s. \quad (9.6)$$

где

$$A(n) = \sum_{i=0}^{N-1} \frac{s(i)F_s}{2\pi(n-i)F_\Delta} \sin\left(\frac{2\pi(n-i)}{F_s} F_\Delta\right) \cos\left(\frac{2\pi}{F_s} \varphi_c(n, i)\right);$$

$$B(n) = \sum_{i=0}^{N-1} \frac{-s(i)F_s}{2\pi(n-i)F_\Delta} \sin\left(\frac{2\pi(n-i)}{F_s} F_\Delta\right) \sin\left(\frac{2\pi}{F_s} \varphi_c(n, i)\right).$$

$s(n)$ - входной сигнал;

$$\varphi_c(n, i) = \begin{cases} \sum_{j=n}^i F_c(j), & n < i \\ -\sum_{j=i}^n F_c(j), & n > i \\ 0, & n = i \end{cases}.$$

$F_c(j)$ – центр полосы пропускания фильтра (может изменяться с каждым отсчетом); $2F_\Delta$ – ширина полосы пропускания фильтра.

Для локализации периодических компонент в частотной области и определения центральной частоты полосы пропускания фильтров в синусоидальном анализаторе используется итеративный перерасчет оцениваемых частот. На каждом шаге полоса пропускания каждого фильтра перемещается в соответствии с полученными оценками для того, чтобы поместить пик энергии оцениваемого компонента в центре полосы (рис. 9.3).

На первом шаге весь частотный диапазон анализируемого сигнала разбивается на полосы B_1, \dots, B_h (h – число полос) с центральными частотами $F_C^{B_1}, \dots, F_C^{B_h}$ соответственно. На каждом последующем шаге оценочные мгновенные частоты $f^{B_1}(n_c), \dots, f^{B_h}(n_c)$ определяются в момент времени, соответствующий центральному отсчету окна анализа n_c . Центральные частоты фильтров изменяются в соответствии с ними: $F_C^{B_x} = f^{B_x}(n_c)$.

После того, как все энергетические максимумы локализованы, можно определить требуемые синусоидальные параметры (амплитудная огибающая, мгновенная частота и фаза). Параметры определяются при помощи выражений (9.4)-(9.6). Так как в процессе анализа несколько фильтров могут локализовать один и тот же компонент, с целью исключения дублирования параметров выполняется сравнение центральных частот $F_C^{B_1}, \dots, F_C^{B_h}$.

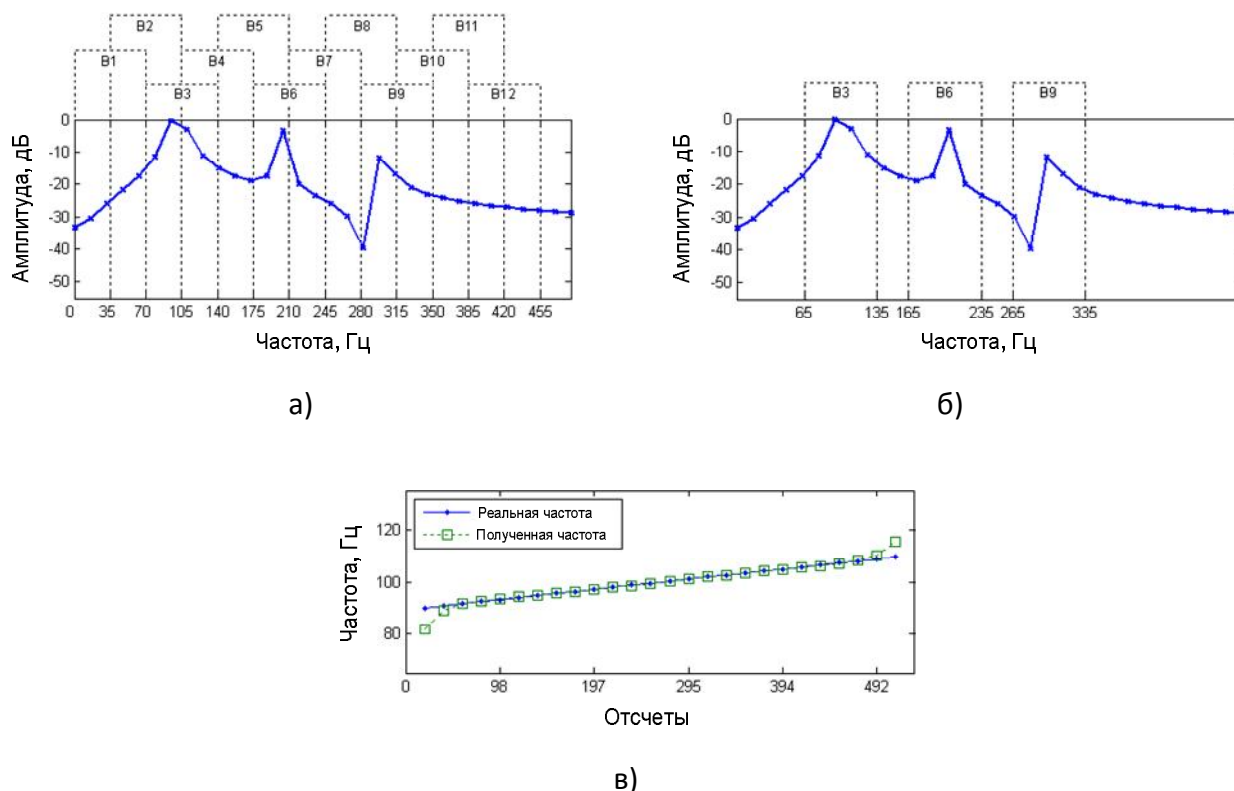


Рис. 9.3 – Кратковременный анализ звукового сигнала с итеративной локализацией периодических компонент ($N = 512$), ($B_1 - B_{12}$) – полосы пропускания фильтров: (а) – первоначальное разбиение частотного диапазона; (б) – локализация компонент после второй итерации; (в) – оценка мгновенной частоты в полосе B_3 .

Чтобы отбросить короткие и нестабильные компоненты, которые являются составляющими транзиентов или шума, выполняется слежение за полученными синусоидальными параметрами. Для выделения только достаточно продолжительных и стабильных значений сравниваются частоты и амплитуды компонентов соседних фреймов.

Данная схема анализа была использована в составе гибридного кодера для звука и речи [33]. Благодаря выявлению продолжительных и стабильных компонент гармонический анализатор способен обрабатывать непосредственно входной сигнал, поступающий на вход систе-

мы кодирования, без какой-либо предварительной обработки. Ниже приведен алгоритм гармонического анализа звукового сигнала.

Исходные параметры алгоритма гармонического анализа звукового сигнала:

N – число отсчетов анализируемого сигнала;

$s(n)$ – анализируемый сигнал, $0 \leq n \leq N - 1$;

F_s – частота дискретизации анализируемого сигнала;

N_{frm} – число отсчетов, соответствующих длине одного фрейма анализа;

N_{off} – число отсчетов, соответствующих смещению между соседними фреймами;

H – число полос, на которые разбивается частотный диапазон сигнала;

F_{Δ} – ширина полосы фильтров анализа деленная на два $F_{\Delta} = \frac{F_s}{2(H+1)}$;

$F_{C,0}^0, \dots, F_{C,H-1}^0$ – первоначальный набор центральных частот фильтров анализа, разбивающих частотный диапазон сигнала $[0, F_s/2]$; $F_{C,h}^0 = F_{\Delta}(h + 1)$, $0 \leq h \leq H - 1$;

I – число итераций, используемых для локализации энергетических максимумов спектра;

$W(m)$ – окно Хэмминга длиной в один фрейм анализа $0 \leq m \leq N_{frm} - 1$;

F_{thr} – максимально допустимая разница между мгновенными частотами одного гармонического компонента в смежных фреймах;

MAG_{thr} – максимально допустимая разница между мгновенными амплитудами одного гармонического компонента в смежных фреймах;

L_{min} – минимальная продолжительность гармонических компонентов.

Алгоритм гармонического анализа звукового сигнала:

Шаг 1. Сегментирование анализируемого сигнала $s(n)$ на фреймы $s^j(m)$, длиной N_{frm} со смещением N_{off} ; умножение полученных фреймов на оконную функцию $s^j(m) = s^j(m)W(m)$.

Шаг 2. Определение мгновенных частот $f_0^j(m_c), \dots, f_{H-1}^j(m_c)$ для фреймов $s^j(m)$ в моменты времени, соответствующие середине фреймов $m_c = N_{frm}/2$ в каждой полосе $[F_{C,0}^0 - F_{\Delta}, F_{C,0}^0 + F_{\Delta}], \dots, [F_{C,H-1}^0 - F_{\Delta}, F_{C,H-1}^0 + F_{\Delta}]$ при помощи соотношения (9.6); установка текущего номера итерации $i = 1$.

Шаг 3. Переопределение мгновенных частот $f_0^j(m_c), \dots, f_{H-1}^j(m_c)$ для фреймов $s^j(m)$ в моменты времени, соответствующие середине фреймов $m_c = N_{frm}/2$ и новому набору полос пропускания $[f_0^j(m_c) - F_{\Delta}, f_0^j(m_c) + F_{\Delta}], \dots, [f_{H-1}^j(m_c) - F_{\Delta}, f_{H-1}^j(m_c) + F_{\Delta}]$ при помощи соотношения (9.6); увеличение текущего номера итерации $i = i + 1$. Шаг 3 выполняется в цикле до тех пор, пока $i < I$.

Шаг 4. Вычисление наборов гармонических параметров (амплитуды $MAG_0^j(m_c), \dots, MAG_{H-1}^j(m_c)$, частоты $f_0^j(m_c), \dots, f_{H-1}^j(m_c)$ и фазы $\varphi_0^j(m_c), \dots, \varphi_{H-1}^j(m_c)$), соответствующих локализованным максимумам спектра частотными полосами $[f_0^j(m_c) - F_{\Delta}, f_0^j(m_c) + F_{\Delta}], \dots, [f_{H-1}^j(m_c) - F_{\Delta}, f_{H-1}^j(m_c) + F_{\Delta}]$ при помощи соотношений (9.4)-(9.6).

Шаг 5. Сравнение полученных параметров отдельно для каждого фрейма $s^j(m)$. Сравниваются все возможные пары полученных частот $f_a^j(m_c), f_b^j(m_c)$, $a, b \in [0, H - 1]$, $a \neq b$. В

случае, если $|f_a^j(m_c) - f_b^j(m_c)| < 2F_\Delta$, тройка параметров $MAG_x^j(m_c)$, $f_x^j(m_c)$, $\varphi_x^j(m_c)$ отображается. $x = a$, если $MAG_a^j(m_c) < MAG_b^j(m_c)$ и $x = b$ в противном случае.

Шаг 6. Сравнение полученных параметров смежных фреймов $s^j(m)$ и $s^{j+1}(m)$. Сравниваются все возможные пары полученных частот $f_a^j(m_c)$, $f_b^{j+1}(m_c)$ и амплитуд $MAG_a^j(m_c)$, $MAG_b^{j+1}(m_c)$, $a, b \in [0, H - 1]$. В случае, если $|f_a^j(m_c) - f_b^{j+1}(m_c)| \leq F_{thr}$ и $|MAG_a^j(m_c) - MAG_b^{j+1}(m_c)| \leq MAG_{thr}$, считается, что тройки параметров $MAG_a^j(m_c)$, $f_a^j(m_c)$, $\varphi_a^j(m_c)$ и $MAG_b^{j+1}(m_c)$, $f_b^{j+1}(m_c)$, $\varphi_b^{j+1}(m_c)$ относятся к одному и тому же гармоническому компоненту. Их длина увеличивается на 1.

Шаг 7. Сравнение полученных параметров смежных фреймов $s^j(m)$ и $s^{j+1}(m)$. Сравниваются все возможные пары полученных частот $f_a^j(m_c)$, $f_b^{j+1}(m_c)$ и амплитуд $MAG_a^j(m_c)$, $MAG_b^{j+1}(m_c)$, $a, b \in [0, H - 1]$. В случае, если $|f_a^j(m_c) - f_b^{j+1}(m_c)| \leq F_{thr}$ и $|MAG_a^j(m_c) - MAG_b^{j+1}(m_c)| \leq MAG_{thr}$, считается, что тройки параметров $MAG_a^j(m_c)$, $f_a^j(m_c)$, $\varphi_a^j(m_c)$ и $MAG_b^{j+1}(m_c)$, $f_b^{j+1}(m_c)$, $\varphi_b^{j+1}(m_c)$ относятся к одному и тому же гармоническому компоненту. Длина данного компонента увеличивается на 1.

Шаг 8. Исключение троек параметров $MAG_h^j(m_c)$, $f_h^j(m_c)$, $\varphi_h^j(m_c)$, которые относятся к гармоническим компонентам с длиной меньше чем L_{min} .

Шаг 9. Синтез периодического сигнала по формуле (9.1), используя оставшиеся тройки параметров $MAG_h^j(m_c)$, $f_h^j(m_c)$, $\varphi_h^j(m_c)$. Вычитание синтезированного сигнала из анализируемого для получения остатка.

9.4.2 Гармонический анализ речевого сигнала

Специфика речевого сигнала позволяет использовать специальную схему гармонического анализа для достижения максимальной точности оцениваемых параметров (рис. 9.4).

Вышеприведенная схема анализа позволяет оценить частоту основного тона и классифицировать текущий фрейм как вокализованный, либо невокализованный. Вокализованные фреймы анализируются с учетом полученной частоты основного тона при помощи частотно-модулированных фильтров. Частота основного тона уточняется для правильной локализации гармоник высокого порядка.

Частотно-модулированные фильтры хорошо применимы к анализу речевых сигналов, так как вокализованная речь имеет гармоническую структуру. Процедура оценки гармонических параметров состоит из двух основных этапов:

- оценка частоты основного тона и классификация фрейма как вокализованного или невокализованного;
- оценка гармонических параметров с уточнением частоты основного тона.

Частота основного тона определяется предварительно, чтобы найти приблизительные частотные траектории гармоник, необходимые для синтеза ЧМ фильтров.

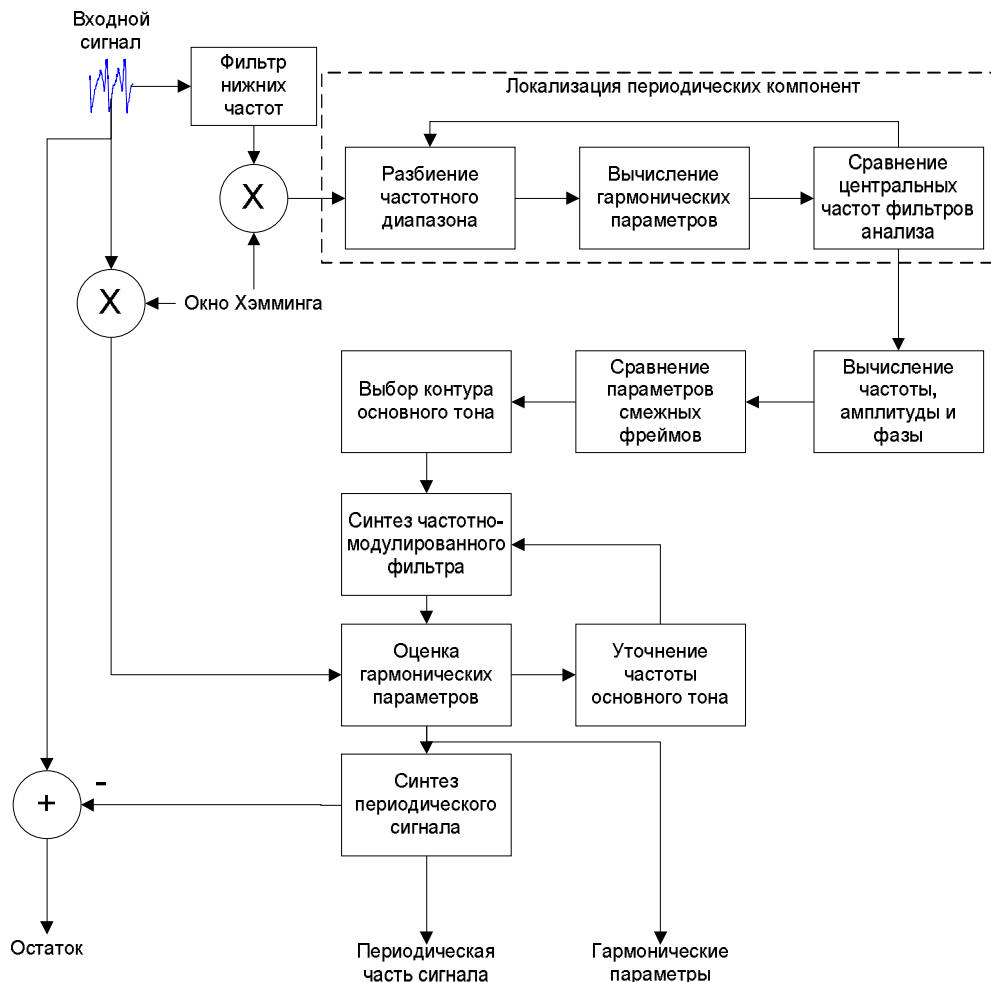


Рис. 9.4 – Схема гармонического анализатора речи

Задача оценки частоты основного тона сводится к нахождению тонального компонента с минимальной частотой и энергией выше определенного амплитудного порога. Диапазон изменения частоты основного тона в речевых приложениях обычно находится в интервале от 60 до 420 Гц. Частоты периодических компонент речевого сигнала в этом диапазоне изменяются довольно медленно, что делает возможным их оценку с использованием стационарных фильтров. Поэтому для определения контуров частоты основного тона целесообразно применение способа оценки мгновенных синусоидальных параметров с последующим выбором из их числа наиболее подходящих путем сравнения. Для того чтобы облегчить процесс вычисления алгоритма оценки, можно воспользоваться ограниченностью частотного интервала поиска и применить предварительную низкочастотную фильтрацию с децимацией. В том случае, когда не удастся выделить продолжительный синусоидальный компонент, соответствующие фреймы сигнала классифицируются как невокализованные.

ЧМ фильтры после оценки частоты основного тона можно использовать для определения гармонических параметров вокализованных фреймов сигнала. Необходимые траектории центральных частот полос пропускания фильтров $F_c(n)$ вычисляются как мгновенная частота основного тона, умноженная на номер k соответствующей гармоники $F_c^k(n) = k f_0(n)$. Процедура оценки проводится последовательно, начиная с первой гармоники и заканчивая послед-

ней. После оценки каждой следующей гармоники контур частоты основного тона уточняется с учетом полученных параметров по следующей формуле перерасчета:

$$f_0(n) = \sum_{i=0}^k \frac{f_i(n) \text{MAG}_i(n)}{(i+1) \sum_{j=0}^k \text{MAG}_j(n)}. \quad (9.7)$$

Таким образом, при оценке гармоник высокого порядка значения частоты основного тона становятся более точными, что позволяет правильно синтезировать частотно-модулированный фильтр анализа. Параметры гармоник вычисляются по соответствующим формулам (9.4)-(9.6).

Ниже приведен алгоритм оценки гармонических параметров речевого сигнала.

Исходные параметры алгоритма гармонического анализа речевого сигнала:

N – число отсчетов анализируемого сигнала;

$s(n)$ – анализируемый сигнал, $0 \leq n \leq N - 1$;

F_s – частота дискретизации анализируемого сигнала;

N_{frm} – число отсчетов, соответствующих длине одного фрейма анализа при оценке контура частоты основного тона;

N_{off} – число отсчетов, соответствующих смещению между соседними фреймами при оценке контура частоты основного тона;

\tilde{N}_{frm} – число отсчетов, соответствующих длине одного фрейма анализа при оценке гармонических параметров;

\tilde{N}_{off} – число отсчетов, соответствующих смещению между соседними фреймами при оценке гармонических параметров;

H – число полос, на которые разбивается частотный диапазон сигнала при оценке контура частоты основного тона;

F_0^{max} – максимально допустимая частота основного тона;

F_Δ – ширина полосы фильтров анализа, используемая при оценке контура частоты основного тона, деленная на два $F_\Delta = F_0^{max} / (H + 1)$;

\tilde{F}_Δ – ширина полосы фильтров анализа, используемая при оценке гармонических параметров, деленная на два (определяется в соответствии);

$F_{C,0}^0, \dots, F_{C,H-1}^0$ – первоначальный набор центральных частот фильтров анализа, разбивающих частотный диапазон сигнала $[0, F_s/2]$; $F_{C,h}^0 = F_\Delta(h + 1)$, $0 \leq h \leq H - 1$;

I – число итераций, используемых для локализации энергетических максимумов спектра;

$W(m)$ – окно Хэмминга длиной в один фрейм анализа, используемое для оценки контура частоты основного тона $0 \leq m \leq N_{frm} - 1$;

$\tilde{W}(m)$ – окно Хэмминга длиной в один фрейм анализа, используемое для оценки гармонических параметров $0 \leq m \leq \tilde{N}_{frm} - 1$;

F_{thr} – максимально допустимая разница между мгновенными частотами одного гармонического компонента в смежных фреймах;

MAG_{thr} – максимально допустимая разница между мгновенными амплитудами одного гармонического компонента в смежных фреймах;

L_{min} – минимальная продолжительность гармонических компонентов;

MAG_0^{min} – минимально допустимая амплитуда основного тона.

Алгоритм гармонического анализа речевого сигнала:

Шаг 1. Фильтрация сигнала $s(n)$ цифровым фильтром с полосой пропускания $[0, F_0^{max}]$ и последующая децимация сигнала с шагом $F_s/(2F_0^{max})$. В результате формируется низкочастотный сигнал $\bar{s}(n)$, который используется для предварительной оценки частоты основного тона.

Шаг 2. Выполнение для $\bar{s}(n)$ шагов 1-8 алгоритма гармонического анализа звукового сигнала, в результате чего формируются тройки гармонических параметров $MAG_h^j(m_c)$, $f_h^j(m_c)$, $\varphi_h^j(m_c)$, которые соответствуют амплитуде, частоте и фазе гармонических компонентов фрейма с номером j и полосе пропускания h .

Шаг 3. Поиск для каждого фрейма с номером j тройки гармонических параметров $MAG_h^j(m_c)$, $f_h^j(m_c)$, $\varphi_h^j(m_c)$ с минимальной частотой $f_h^j(m_c)$ и амплитудой не менее, чем MAG_0^{min} : $MAG_h^j(m_c) \geq MAG_0^{min}$. В случае нахождения такой тройки параметров фрейм классифицируется как вокализованный, в противном случае фрейм классифицируется как невокализованный. Из всех гармонических параметров фрейма j сохраняется только частота найденной тройки параметров $f_h^j(m_c)$, все остальные параметры отбрасываются.

Шаг 4. Формирование предварительного контура частоты основного тона $f_0(n)$ путем линейной интерполяции полученных значений минимальной частоты $f_h^j(m_c)$ в моменты времени, соответствующие отсчетам исходного сигнала $s(n)$.

Шаг 5. Сегментирование исходного сигнала $s(n)$ и контура основного тона $f_0(n)$ на фреймы длиной \tilde{N}_{off} со смещением \tilde{N}_{off} , в результате которого формируются фреймы $s^j(m)$ и $f_0^j(m)$ соответственно. Умножение фреймов $s^j(m)$ на оконную функцию $s^j(m) = s^j(m)\tilde{W}(m)$. Расчет числа анализируемых гармоник для каждого фрейма $K^j = \min_m(F_s/f_0^j(m))$. Установка текущего номера гармоники $k = 0$.

Шаг 6. Формирование контура центральных частот фильтров анализа $F_c^j(m)$ для гармоники с номером k $F_c^j(m) = (k + 1)f_0^j(m)$. Для всех фреймов с номером j , у которых число анализируемых гармоник больше номера текущей гармоники $K^j > k$, при помощи соотношения (9.4)-(9.6) определяются тройки гармонических параметров $MAG_k^j(m_c)$, $f_k^j(m_c)$, $\varphi_k^j(m_c)$ в моменты времени, соответствующие середине фреймов $m_c = N_{frm}/2$. Для всех фреймов с номером j , у которых $K^j \leq k$, гармонические параметры считаются равными нулю $MAG_k^j(m_c) = 0$, $f_k^j(m_c) = 0$, $\varphi_k^j(m_c) = 0$.

Шаг 7. Уточнение значения контура основного тона $f_0^j(m_c)$ в моменты времени m_c для каждого фрейма при помощи соотношения (9.7). Формирование нового контура частоты основного тона $f_0(n)$ путем линейной интерполяции полученных значений $f_0^j(m_c)$ в моменты времени, соответствующие отсчетам исходного сигнала $s(n)$. Текущий номер гармоники увеличивается на единицу $k = k + 1$.

Шаг 8. Сегментирование уточненного контура основного тона $f_0(n)$ на фреймы длиной \tilde{N}_{off} со смещением \tilde{N}_{off} . Формируются новые фреймы $f_0^j(m)$. Если существует такой номер фрейма j , для которого $K^j > k$, то осуществляется переход к шагу 6.

Шаг 9. Синтез периодического сигнала по формуле (9.1), используя оставшиеся тройки параметров $MAG_{0\dots K-1}^j(m_c)$, $f_{0\dots K-1}^j(m_c)$, $\varphi_{0\dots K-1}^j(m_c)$. Вычитание синтезированного сигнала из анализируемого для получения остатка.

9.5 Моделирование переходных составляющих сигнала методом согласованной подгонки

Алгоритмы согласованной подгонки используется в некоторых системах параметрического кодирования для компактного представления переходной составляющей различных сигналов [34,35]. Основными задачами согласованной подгонки являются: поиск метода классификации, выбор наиболее значимых компонентов сигнала, а так же определение базисной функции для обеспечения эффективного описания входного сигнала с минимальной ошибкой.

Процесс оптимизации согласованной подгонки может быть основан на знании психоакустических свойств восприятия звуков человеком. Это позволяет адаптировать размер словаря в соответствии с особенностями человеческого слуха. Для того чтобы элементы словаря оптимальным образом поставить в соответствие отдельным сегментам сигнала (с точки зрения величины вносимых искажений) используется адаптивный психоакустический критерий. Такого рода подходы успешно применяются для моделирования сигнала затухающими синусоидами и вейвлетами [15,33,36].

9.5.1 Согласованная подгонка с динамически оптимизируемым словарем на основе психоакустики

Пусть $\{\varphi_n(t): n \in Z\}$ является множеством вейвлетных функций и $E \in \{(l, n): 0 \leq l \leq L, 0 \leq n \leq 2^l\}$ – узлы древовидной структуры. Тогда интервал $[0,1)$ разделяется на двухэлементные интервалы $I_{l,n} = [n2^{-l}, (n+1)2^{-l}]$, которые соответствуют определенному множеству узлов $(l, n) \in E$. В частности $\{\varphi_{l,n,k}(t): (l, n) \in E, k \in Z\}$, где $\varphi_{l,n,k}(t) \triangleq 2^{-l/2}X_n(2^{-l}t - k)$ базовая форма в пространстве сигнала $span\{\varphi_0(t - k): k \in Z\}$. Каждый узел $(l, n) \in E$ дерева соответствует своей частотной полосе. В соответствии с двухэлементной древовидной структурой сигнал разделяется на полосы, которые близки к критическим [37]: $(l, n) \in E_{CB}$, $l = \{0, b\}$, где E_{CB} описывает ограниченную древовидную структуру вейвлет-пакета, b является максимальным номером уровней вейвлет-декомпозиции и зависит от частотного диапазона. Так, для обработки звуковых сигналов b равняется 8. В соответствии с E_{CB} частотная полоса $[0-22.05 \text{ kHz}]$ делится на 25 полос. Коренной узел $(l, n) = (0,0)$ данного дерева соответствует всему частотному диапазону сигнала.

Общий алгоритм может быть описан, как аппроксимация анализируемого сигнала $x(n)$ линейным пространством с атомами g_γ , выбираемыми из словаря вейвлет-пакета D [35]. Каждый вектор $g_\gamma \in D$ имеет индекс $\gamma = (l, n, k)$, причем $0 < l < \log_2(N)$, $0 < n < 2^l$, $0 < k <$

$2^{-l}N$, где N длина фрейма сигнала. Данные вектора имеют такие же свойства частотно-временной локализации, как и дискретная оконная функция, расширенная на 2^l и имеющая центр в $2^l(k + 1/2)$.

Метод моделирования переходных сегментов на основе согласованной подгонки с динамически оптимизируемым словарем при помощи психоакустического критерия состоит из двух основных этапов. Первый этап представляет собой пофреймную оптимизацию вейвлет пакета, основанную на весовой функции энтропии входного сигнала $x(n)$ [37]; второй этап моделирования реализуется с помощью алгоритма согласованной подгонки с перцептуальной оценкой [38,39]. Результатами первого этапа являются: дерево вейвлет пакета E_j , оптимизированное для текущего фрейма входного сигнала $x(n)$; порог маскирования $T_{l,n}$, и порог временного маскера $F_{l,n}$ в узлах древовидной структуры вейвлет-пакета E_j [37]; скалограмма аудиторного возбуждения для входного сигнала $x(n)$, вычисленная с использованием $T_{l,n}$ и $F_{l,n}$ для всех узлов [38]. Таким образом, на первом этапе моделирования переходных фреймов на основе согласованной подгонки (см. рис. 9.5), входной сигнал $x(n)$ разделяется на субполосы при помощи банка фильтров, на основе психоакустически адаптивного дерева вейвлет пакета для данного фрейма входного сигнала. Каждый вейвлет коэффициент соответствует скалярному произведению входного сигнала и атома g_γ из словаря. Наиболее значимые компоненты можно определить при помощи классификации выбранных, перцептуально значимых коэффициентов вейвлет пакета, которые выбираются следующим образом. Каждый новый добавляемый коэффициент выбирается таким образом, чтобы обеспечить максимальный прирост в согласованной подгонке между скалограммами аудиторного возбуждения $G_{l,n,k}$, соответствующими исходному и полученному сигналам [39]. Скалограммы сигналов строятся с использованием маскирующих порогов $T_{l,n}$ в вейвлет области. Выбирается коэффициент с максимальным по модулю значением. Вклад вектора $\alpha_\gamma^m \cdot g_\gamma^m(n)$ затем вычитается из сигнала $x(n)$ и процесс повторяется для остатка $r(n)$. После итерации m остаток $r^m(n)$ представляет собой:

$$r^m(n) = \begin{cases} x(n) & m = 0 \\ r^{m+1}(n) + \alpha_\gamma^m \cdot g_\gamma^m(n) & m \neq 0 \end{cases} \quad (9.8)$$

где α_γ^m - вес, соответствующий вектор оптимума $g_\gamma^m(n)$ итерации m ; γ - индекс словаря вейвлет пакета итерации m .

Вектор оптимума – это вектор максимального скалярного произведения и сигнала остатка $\langle r^m, g_\gamma^m \rangle$. Для того, чтобы найти компактное представление сигнала, каждый коэффициент вейвлет пакета, имеющий наибольший вес возбуждения, добавляется в параметрическое описание. Вес возбуждения соответствует разнице между исходной и моделируемой скалограммами коэффициентов возбуждения.

Алгоритм согласованной подгонки может быть реализован так, как показано ниже.

Исходные данные алгоритма согласованной подгонки:

E_j – оптимизированная древовидная структура вейвлет пакета для сигнала $x(n)$;

$T_{l,n}$ – порог маскирования;

$F_{l,n}$ – временный маскер в узлах E_j ;

$G_{l,n,k}$ – скалограмма аудиторного возбуждения соответствующая сигналу $x(n)$.

Алгоритм согласованной подгонки:

Шаг 1. Установить номер итерации $m = 0$.

Шаг 2. Разместить $G_{l,n,k}$ и установить $G_{l,n,k} = 0$ для всех l, n, k в соответствии с древовидной структурой вейвлет пакета E_j .

Шаг 3. Вычислить $PE_{l,n}^m$ для всех узлов (l, n) , используя $T_{l,n}$;

если $PE_{l,n}^m == 0 \forall (l, n, k) \in E_j$ СТОП;

если $PE_{l,n}^m == 0$, то $X_{l,n,k}^m = 0$ для $k = \{0, K_{l,n} - 1\}$ узла (l, n) .

Шаг 4. Выбрать из $X_{l,n,k}^m$ значимые коэффициенты $X_{l,n,k}^{*m}$, имеющие наибольший вес возбуждения.

Шаг 5. Создать скалограмму аудиторного возбуждения, соответствующую моделируемому сигналу, используя $T_{l,n}$ и $F_{l,n}^{m-1}$ для выполненной итерации и каждого нового значимого коэффициента $X_{l,n,k}^{*m}$.

Шаг 6. Выбрать вес $\alpha_{l,n,k}^m = X_{l,n,k}^{*m}$, который улучшает соответствие между скалограммами (исходной и моделируемой).

Шаг 7. Получить позицию коэффициента вейвлет-пакета: $l^* = l, n^* = n, k^* = k$.

Шаг 8. Установить 1 на позиции (l^*, n^*, k^*) : $G_{l^*,n^*,k^*} = 1$.

Шаг 9. Синтезировать атом $g^m(n)$ из G_{l^*,n^*,k^*} , используя обратный вейвлет-пакет с древовидной структурой E_j , соответствующей словарю вейвлет-пакета.

Шаг 10. Вычислить остаточный сигнал $r^m(n)$ из $g^m(n)$ и $\alpha_{l,n,k}^m$ в соответствии с (9.8).

Шаг 11. Применить оптимизированный на основе текущего фрейма вейвлет-пакет с древовидной структурой E_j к остаточному сигналу $r^m(n)$.

Шаг 12. Увеличить номер итерации $m = m + 1$.

Шаг 13. перейти к Шагу 2.

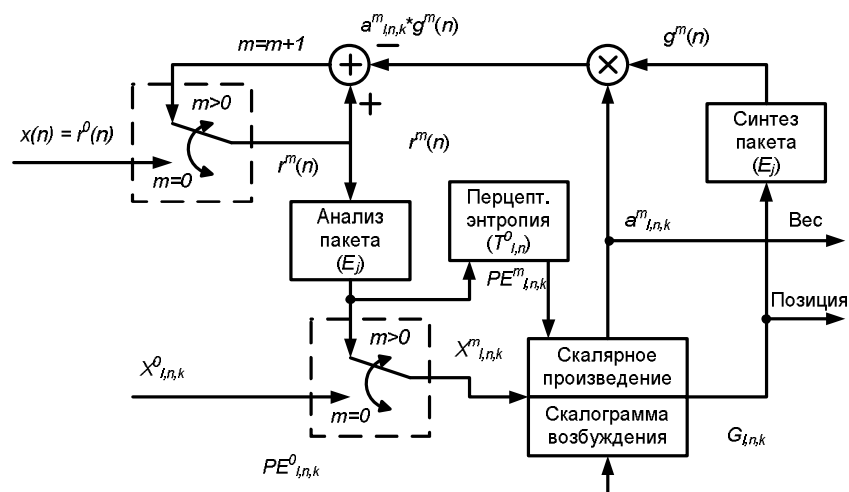


Рис. 9.5 – Блок-схема согласованной подгонки с динамически оптимизируемым словарем на основе психоакустического критерия

Преимуществом алгоритма является минимизация перцептуального искажения, определяемого при помощи перцептуально оптимизированного частотно-временного плана соответствующей декомпозиции вейвлет пакета для выбора оптимального атома для каждой ите-

рации подгонки. Более того, существует точный психоакустический критерий остановки описанной процедуры. Число итерации алгоритма согласованной подгонки для анализируемого фрейма определяется числом перцептуально важных коэффициентов вейвлет пакета.

Сравнительная сходимость трех различных методов согласованной подгонки показана на рис. 9.6. Описанная процедура обеспечивает наименьшее среднеквадратичное отклонение, по сравнению с методами на основе затухающих синусоид и вейвлет пакетов с избыточным словарем.

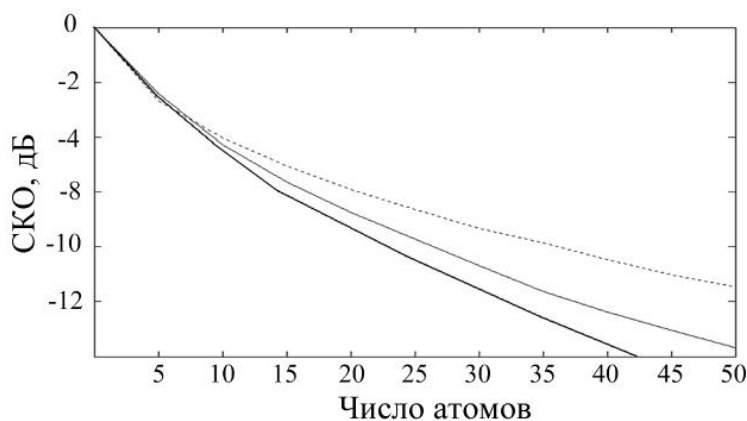


Рис. 9.6 – Среднеквадратичное отклонение: затухающие синусоиды (пунктирная линия); вейвлет пакет с избыточным словарем (тонкая сплошная линия); вейвлет пакет, оптимизированный для каждого фрейма на основе психоакустики (толстая сплошная линия)

9.5.2 Определение переходных фреймов

Способ определения переходных фреймов основан на предположении, что энергия остаточного сигнала после синусоидального моделирования (переходная + шумовая составляющие) резко изменяется при наличии выраженной переходной составляющей [35]. Изменения энергии могут быть общими, либо представлять собой перераспределение между разными частотными полосами. Остаточный сигнал переводится в область вейвлетов, используя декомпозицию при помощи вейвлет пакета второго уровня. Алгоритм вычисляет энергию вейвлет коэффициентов в каждой полосе. Затем энергия каждой полосы фрейма i делится на энергию соседних фреймов $(i - 1)$ и $(i + 1)$, после чего полученный коэффициент сравнивается с порогом. Порог задается в зависимости от синусоидальных амплитуд, полученных на стадии гармонического анализа, для того, чтобы отбросить маскированные переходные составляющие.

9.6 Применение гибридной системы анализа

В данном подразделе приведена сепарация звукового сигнала (звук колокольчиков с частотой дискретизации 44,1 кГц) на синусоидальную, переходную и шумовую составляющие. Каждая составляющая показана в виде волновой формы и спектрограммы для качественной демонстрации, дающей наиболее полное представление о гибридном подходе. На рис. 9.7 приведены все сигналы, возникающие на различных этапах анализа.

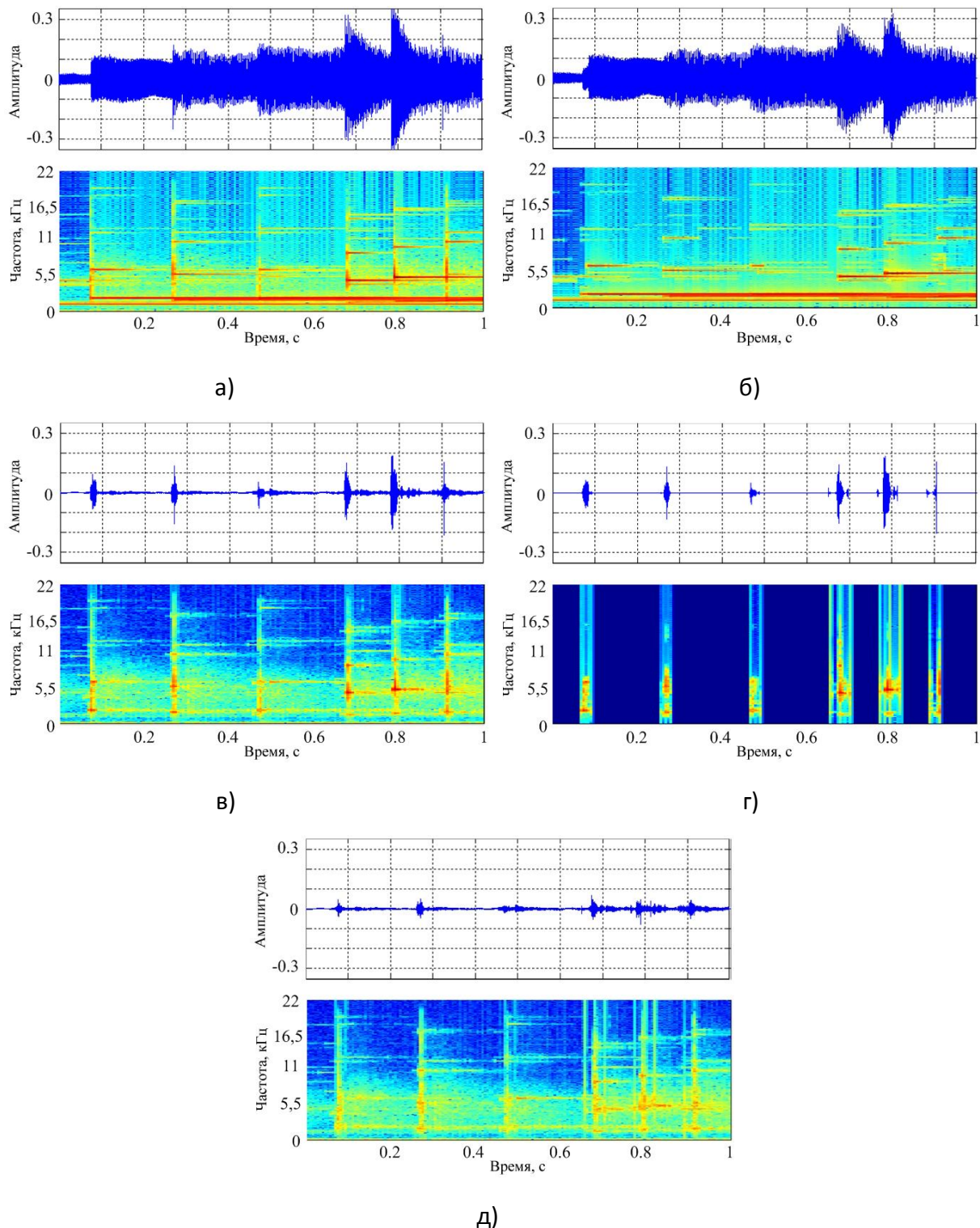


Рис. 9.7 – Сепарация сигнала гибридной системой анализа: а – входной сигнал; б – синтезированная периодическая часть; в – остаток (переходная составляющая + шум); г – переходная часть сигнала (моделирование методом согласованной подгонки); д – остаток (шум)

Синусоидальный анализ выполнялся с разбиением на фреймы длиной 48 мс и смещением 14 мс между соседними фреймами. Использованная ширина полос фильтров анализа составляла 70 Гц, число итераций для локализации периодических компонент – 5. Из рис. 9.7 видно, что входной сигнал (рис. 9.7а) содержит компоненты: гармонические, переходные и шумовые. На выходе гармонического анализатора формируются две различные составляющие

сигнала: периодическая (рис. 9.7б) и остаточная (рис. 9.7в). Синтезированная периодическая часть содержит лишь продолжительные синусоидальные компоненты с высокой локализацией энергии, в то время как шум и переходные фрагменты сигнала находятся в остаточной части. Отношение энергии периодической части к остаточной составляет 14,77 дБ, что свидетельствует о гармонической природе анализируемого сигнала.

Переходные компоненты сигнала, которые были найдены из остатка синусоидального моделирования, обрабатывались вышеизложенным алгоритмом согласованной подгонки на основе перцептуально оптимизированных для каждого фрейма вейвлет пакетов. Входной остаточный сигнал был разделен на фреймы длиной по 1024 отсчета. В эксперименте были использованы фильтры 40-го порядка.

Воссозданные из параметрического описания переходные сегменты сигнала потребовали 20, 23, 18, 32, 36, 25, 27 и 20 атомов соответственно. В результате сложения синусоидальной, переходной и шумовой компонент получится исходный сигнал.

9.7 Заключение

Усовершенствованный гармонический анализ с трекингом параметров позволяет должным образом и без дополнительной обработки разделять входной сигнал на периодическую и остаточную компоненты, оставляя переходные сегменты сигнала в неизменном виде. Использование гармонического анализа в начале схемы параметризации существенно упрощает дальнейшую обработку сигнала (в частности, определение переходных сегментов сигнала). Описанная методология для выбора наиболее значимых вейвлет коэффициентов основывается на максимизации соответствия между аудиторными скалограммами, соответствующими исходному и синтезированному сигналам. Основное достоинство изложенного метода заключается в том, что словарь вейвлет пакета оптимизирован на основе принципов психоакустики отдельно для каждого фрейма сигнала, что значительно понижает общее число коэффициентов, необходимых для достижения заданного уровня перцептуальных искажений.

9.8 Список использованных источников

1. Ali A.M.A., Robust auditory-based speech processing using the average localized synchrony detection, IEEE Transactions on Speech and Audio Processing, 2002, Vol.10, No 5, pp. 279–292.
2. Wang S. and et al., An objective measure for predicting subjective quality of speech coders, IEEE J. on selected areas in communications, 1992, vol. 10, No. 5, pp. 819–829.
3. Bielawski A., Petrovsky A.A., Speech enhancement system for hands-free telephone based on the psychoacoustically motivated filter bank with allpass frequency transformation, Speech communication and technology: proceedings of 6-th europ. conf. (EUROSPEECH'99), Budapest, Hungary, September 5–9, 1999, Budapest, 1999, pp. 2555–2558.
4. Delgutte B., Speech coding in the auditory nerve: Voiceless fricative consonants, J. Acoust. Soc. Am., 1984, Vol. 75, pp. 887–896.

-
5. Ghitza O., Adequacy of auditory models to predict internal human representation of speech sounds, *J. Acoust. Soc. Am.*, 1993, Vol. 93, No. 4, pp. 2160–2171.
 6. Ghitza O., Auditory Nerve Representation Criteria for Speech Analysis/Synthesis, *IEEE Transactions on Speech and Audio Processing*, 1987, Vol. ASSP-35, No. 6, pp. 736–740.
 7. Parfieniuk M., Petrovsky A., Warped DFT as the basis for psychoacoustical model, *Acoustics, speech and signal processing: proceedings of IEEE int. conf. (ICASSP-2004)*, Montreal, Canada, May 17–21, Montreal, 2004, Vol. 4, pp. 185–188.
 8. Zwicker E., Fastl H., *Psychoacoustics: facts and models*, Berlin: Springer-Verlag, 1990, 354 p.
 9. Zwislocki J., Analysis of some auditory characteristics // *Handbook of Mathematical Psychology*, New York: John Wiley and Sons Inc., 1965, 267 p.
 10. George E.B., Smith M.J.T., Speech Analysis/Synthesis and Modification Using an Analysis-by-Synthesis/Overlap-Add Sinusoidal Model, *IEEE Trans. on Speech and Audio Process*, 1997, Vol.5, No. 5. pp. 389-406.
 11. Levine S., Smith J., A sines+transients+noise audio representation for data compression and time/pitch scale modifications, *Signal processing: proceedings of 105th AES convention*, San Francisco, USA, 1998, San Francisco, 1998, Preprint № 4781, 21 pages.
 12. Dutoit T., *An Introduction to Text-to-speech Synthesis*, The Netherlands: Kluwer Academic Publishers, 1997, 326 p.
 13. Laroche J., Stylianou Y., Moulines E., HNS: Speech modification based on a harmonic+noise model, *Acoustic, Speech, and Signal Processing: proceedings of IEEE International conference ICASSP-93*, Minneapolis, USA, 27-30 April 1993, pp. 550-553.
 14. Accardi A.J., Cox R.V., A modular approach to speech enhancement with an application to speech coding, *Acoustics, speech, and signal processing: proceedings of IEEE int. conf.*, Phoenix, Arizona, USA, March 15–19, 1999, pp. 201–204.
 15. Petrovsky A., Azarov E., Petrovsky A., Combining advanced sinusoidal and waveform matching models for parametric audio/speech coding, *USIPCO 2009: proc. of the 17th European Signal Processing Conference*, Glasgow., 2009, pp. 436-440.
 16. Das A., Gersho A., Low-rate multimode multiband spectral coding of speech, *Int. J. of speech technology*, 1999, Vol. 2, pp. 317–327.
 17. George E.B., Smith M.J.T., Perceptual considerations in a low bit rate sinusoidal vocoder, *Computers and communications: proceedings of int. conf. (IPCCC'90)*, Phoenix, USA, March 21–23, 1990, P. 268–275.
 18. Ghitza O., Auditory Models and Human Performance in Tasks Related to Speech Coding and Speech Recognition, *IEEE Transactions on Speech and Audio Processing*, 1994, Vol. 2, No. 1., pp. 115–132.

-
19. McAulay R.J., Quatieri T.F., Computationally Efficient Sine-wave Synthesis and Its Application to Sinusoidal Transform Coding, *Acoustics, speech and signal processing: proceedings of IEEE int. conf.*, New York, USA, April 11–14, 1988, pp. 370-373.
 20. McAulay R.J., Quatieri T.F., Low-rate speech coding based on the sinusoidal model, *Advances in Speech Signal Processing / Eds. S. Furui, M.M. Sondhi*, New York: Marcel Dekker, 1992, pp. 165–208.
 21. Moorer J., The use of the phase vocoder in computer music applications, *J. of the Audio Engineering Society*, January-February 1978, Vol. 26, pp. 42–45.
 22. McAulay, R.J. Speech analysis/synthesis based on a sinusoidal representation / R.J. McAulay, T.F. Quatieri // *IEEE Trans. on Acoust., Speech and Signal Processing*, 1986, Vol. ASSP-34, pp. 744–754.
 23. McAulay R.J., Quatieri T.F., Sinusoidal coding, *Speech Coding and Synthesis / Eds. W.B. Kleijn, K.K. Paliwal*, Amsterdam: Elsevier, 1995, pp. 121–173.
 24. Bailly G., Accurate estimation of sinusoidal parameters in an harmonic + noise model for speech synthesis, *Speech communication and technology: proceedings of 6-th europ. conf. (EUROSPEECH'99)*, Budapest, Hungary, September 5–9, 1999, pp. 1051-1054.
 25. Eriksson J.L., Robert A., The Representation of Pure Tones and Noise in a Model of Cochlear Nucleus Neurons, *J. Acoust. Soc. Am.*, 1999, Vol. 106, No. 4, pp. 1865–1879.
 26. Serra X., Musical Sound Modeling with Sinusoids plus Noise, *Musical Signal Processing / Eds. C. Roads, S. Pope, A. Piccilli, G. De Poli*, Lisse: Swets & Zeitlinger Publishers, 1997, pp. 91-122.
 27. Painter T., Spanias A., Sinusoidal analysis-synthesis of audio using perceptual criteria, *EURASIP Journal on Applied Signal Processing*, 2003, № 1, pp. 15-20.
 28. Almeida L.B., Silva F.M., Variable-frequency synthesis: An improved harmonic coding scheme, *Acoustics, speech and signal processing: proceedings of IEEE int. conf. (ICASSP-84)*, San Diego, USA, March 19–21, 1984, pp. 27–51.
 29. Virette D., and et al., Analysis of background noise reduction techniques for robust speech coding, *Signal processing: proceedings of 11-th europ. conf. (EUSIPCO'2002)*, Toulouse, France, 3–6 September, 2002, Vol. 3, pp. 297–300.
 30. Edler B., Purnhagen H., Parametric Audio Coding, *Signal processing: proceedings of 5-th int. conf. (ICSP'2000)*, Beijing, China, August 21–25, 2000, pp. 21–24.
 31. McAulay R.J., Quatieri T.F., Pitch estimation and voicing detection based on sinusoidal model, *Acoustics, speech and signal processing: proceedings of IEEE int. conf. (ICASSP'90)*, Albuquerque, New Mexico, USA, April 3–6, 1990, pp. 249–252.
 32. Spanias A.S., Speech coding: a tutorial review, *Proc. IEEE.*, 1994, Vol. 82, No. 10, pp. 1541–1582.

-
33. Petrovsky A., Azarov E., Petrovsky A.A., Harmonic representation and auditory model-based parametric matching and its application in speech/audio analysis, Munich, 2009, 13 pages (Preprint / Signal processing: proceedings of 126-th AES convention; № 7705).
 34. Mallat S., Zhang Z., Matching pursuits with time-frequency dictionaries, IEEE. Transactions on Signal Processing, 1993, Vol. 41, No. 12, pp. 3397-3415.
 35. Vera-Candeas P, Ruiz-Reyes N., Roza-Zurera M., Transient modeling by Matching-Pursuits with a wavelet dictionary for parametric audio coding, IEEE Signal Processing Letters, 2004, Vol. 11, No. 3, pp. 349-352.
 36. Verma T, Meng T., Sinusoidal modeling using frame-based perceptually weighted matching pursuits, Acoustics, speech and signal processing: proceedings of IEEE intern. conf. (ICASSP-99), Phoenix, Arizona, USA, March 15–19, 1999, Vol. 2, pp. 981–984.
 37. Petrovsky A., Krahe A., Petrovsky A., Real-time wavelet packet-based low bit rate audio coding on a dynamic reconfigurable system, Signal processing: proceedings of 114-th AES convention, Netherlands, Amsterdam, 2003, Preprint № 5778, 22 pages.
 38. Petrovsky Al., Petrovsky A., Transitional speech segments modeling by matching pursuit with a dictionary based on the psychoacoustic adaptive WP, Proc. of the 11th Int. Conf. "Speech and Computer" (SPECOM'2006), 25-29 June, St. Petersburg, 2006, pp.380-385.
 39. Petrovsky Al., Petrovsky A., Matching pursuit algorithm with frame-based auditory optimized WP-dictionary for audio transient modeling, "ELEKTRONIKA"-PAN, Warsaw, Vol, 49, №4, 2008, pp. 74-79.

10 КОМПРЕССИЯ РЕЧЕВОГО СИГНАЛА НА ОСНОВЕ СИНУСОИДАЛЬНОЙ МОДЕЛИ С АНТРОПОМОРФИЧЕСКОЙ ОБРАБОТКОЙ

Лихачёв Д.С., Петровский А.А.

АННОТАЦИЯ

Описывается система компрессии речевого сигнала на основе синусоидальной модели с антропоморфической обработкой. Отличительной особенностью данной системы является то, что речь, как на вокализованных, так и на невокализованных участках, представляется в виде суммы синусоидальных компонент. Антропоморфическая обработка заключается в моделировании физиологических процессов в слуховой системе человека и получении информации о степени важности для восприятия человеком частотных составляющих анализируемого сигнала. Это позволяет получить восстановленный речевой сигнал с сохранением основных характеристических особенностей речи при ограниченном числе синусоидальных компонент.

Ключевые слова – антропоморфическая обработка речевого сигнала, модели слуха человека, система компрессии.

СОДЕРЖАНИЕ

10.1	Задача антропоморфической обработки речевого сигнала применительно к системам компрессии на основе синусоидальной модели	211
10.2	Антропоморфическая обработка речевого сигнала на основе моделей слуха человека ...	214
10.2.1	Принципы антропоморфического анализа.....	214
10.2.2	Пассивная кохлеарная модель	216
10.2.3	Активная кохлеарная модель	219
10.2.4	Модель слухового нерва.....	222
10.3	Особенности построения системы компрессии речевого сигнала на основе синусоидальной модели с антропоморфической обработкой ..	225
10.3.1	Кодер.....	225
10.3.2	Декодер.....	227
10.4	Список использованных источников.....	230

10.1 Задача антропоморфической обработки речевого сигнала применительно к системам компрессии на основе синусоидальной модели

Все методы обработки и компрессии сигналов, основанные на синусоидальном представлении, базируются на предположении о том, что любой сигнал можно представить в виде суммы синусоидальных компонент с изменяющимися во времени параметрами: амплитудами,

частотами и фазами [1]. Основными недостатками таких систем является то, что для достижения хорошего качества восстанавливаемого речевого сигнала он должен быть представлен достаточно большим количеством синусоидальных параметров, что непригодно для кодирования речи с высокой степенью компрессии.

При ограничении количества синусоидальных компонент резко деградирует качество речи. Как показали многочисленные работы в данной области, количество необходимых синусоидальных компонент для высококачественного кодирования речи находится в пределах от 40 до 80 штук на отрезок сигнала длительностью 20 мс и не может быть значительно уменьшено при использовании стандартных алгоритмов синусоидального анализа [2-4].

Для решения этой проблемы был предложен ряд речевых кодеров, где тем или иным способом ограничивается объём кодируемой информации [5-7]. В настоящее время интенсивно развиваются синусоидальные кодеры с антропоморфической обработкой речевого сигнала [2,4,8]. В этих кодерах объём кодируемой речевой информации уменьшается за счёт применения различных слуховых моделей человека и психоакустических принципов обработки речевого сигнала. Применение принципа антропоморфической обработки сигнала в данном случае предполагает использование таких алгоритмов обработки информации, когда вычислительный процесс организовывается по “образу и подобию” человека, т.е. применяемые способы и алгоритмы моделируют какие-либо процессы, происходящие в его слуховых и речеобразующих системах. При этом предполагается, что при достаточно точном моделировании используемые системы будут иметь те же полезные свойства, что и их физиологический аналог. Различные алгоритмы антропоморфической обработки в настоящее время с успехом применяются не только для кодирования сигналов, но и в задачах распознавания речи и идентификации диктора [9].

При “классическом” подходе к синусоидальному кодированию речи используется критерий оптимальности процедуры синтеза, при котором восстановленный в декодере речевой сигнал должен быть наиболее близок к оригинальному сигналу только по его форме.

Антропоморфический критерий оптимальности предъявляет совершенно другие требования: максимально близкой к оригиналу должна быть не форма речевого сигнала, а его отображение на уровне системы слуха человека. Целесообразность применения данного критерия заключается в следующем.

Во-первых, из-за особенностей функционирования системы слуха не все частотные составляющие сигнала одинаково важны при восприятии акустической (звуковой) информации человеком [9].

Во-вторых, в речевом сигнале наблюдаются эффекты маскирования [10]. Следовательно, нет необходимости использовать синусоидальные компоненты, которые не будут слышны из-за эффектов маскирования или которыми можно пренебречь из-за их не очень высокой важности для человеческого слуха.

Таким образом, можно утверждать, что в системах компрессии речевой информации на основе синусоидальной модели для уменьшения количества кодируемой информации, не ухудшая при этом качество восстановленной в декодере речи, необходимо использовать специальные критерии отбора наиболее важной для слуха человека информации [4,5,11].

Для отбора наиболее важных с точки зрения человеческого слуха (доминирующих) синусоидальных составляющих необходимо использовать критерии, которые хорошо согласуются с особенностями восприятия человеком звуковой информации. Процесс анализа на основе синусоидальной модели речевого сигнала с антропоморфическим критерием отбора доминирующих частотных составляющих можно представить, как показано на рис. 10.1. В этом случае восстановленная речь генерируется как сумма синусоид с параметрами, определяемыми на доминирующих частотах.

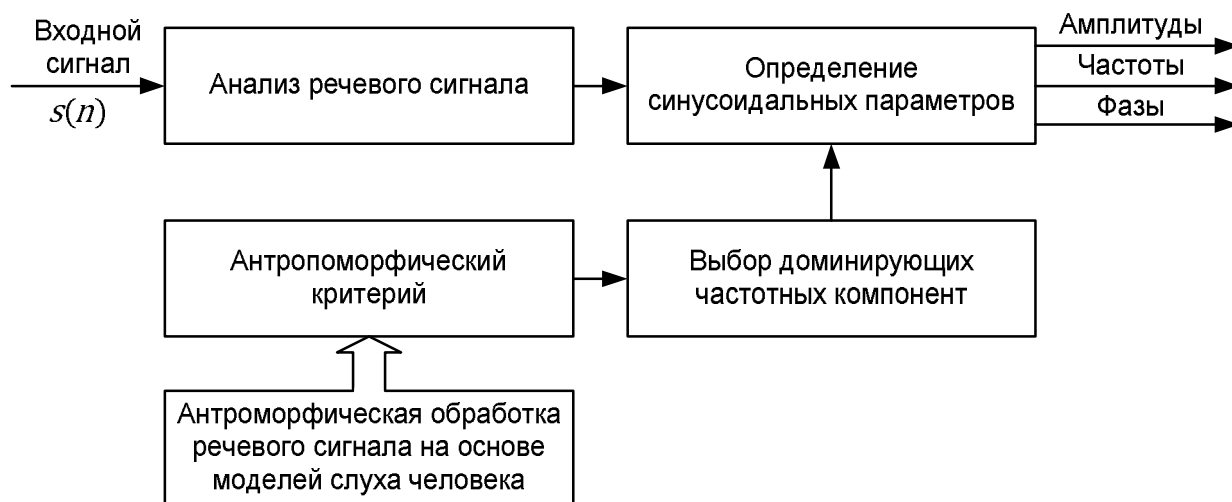


Рис. 10.1 – Обобщённая схема синусоидального анализа с антропоморфической обработкой

Для эффективного использования антропоморфической обработки сигнала применительно к низкоскоростному кодированию речи на основе синусоидальной модели предъявляются следующие требования.

Во-первых, методы антропоморфического анализа речевого сигнала для выделения наиболее важных для слуха человека частотных компонент должны быть достаточно адекватны реальным физиологическим процессам, которые происходят в слуховой системе человека.

Во-вторых, результаты работы алгоритмов антропоморфического анализа должны быть представлены в виде, приемлемом для совместного использования с синусоидальной моделью. Таким образом, при использовании спектрального анализа для определения синусоидальных параметров, в качестве результата антропоморфической обработки должна выступать функция, отображающая важность для человеческого слуха того или иного спектрального компонента.

В-третьих, найденный в результате анализа набор параметров речевого сигнала должен не только обеспечивать информативность и хорошее восприятие человеком восстановленной речи, но и давать возможность достичь высокой степени компрессии за счёт эффективного квантования данных параметров.

10.2 Антропоморфическая обработка речевого сигнала на основе моделей слуха человека

10.2.1 Принципы антропоморфического анализа

В качестве основной задачи для антропоморфической обработки ставится выработка такого критерия, который позволял бы отбирать наиболее важные для восприятия человеком доминирующие частотные компоненты. Поэтому, в данном случае антропоморфический анализ заключается в моделировании физиологических процессов в слуховой системе человека и получении информации о степени важности для восприятия частотных составляющих анализируемого сигнала. При моделировании такой системы возникает проблема дифференциации её элементов по значимости с точки зрения механизма обработки звуковой информации и применимости для задач компрессии. Моделирование “второстепенных” элементов не только сильно увеличивает алгоритмическую и вычислительную сложность, но и усложняет процесс интерпретации полученных результатов. Таким образом, для успешного решения поставленной выше задачи по выделению доминирующих частотных составляющих необходимо не просто смоделировать процессы, происходящие в периферической части слуховой системы человека (главным образом в улитке уха человека), но и проанализировать информацию, которая циркулирует на уровне слухового нерва человека.

Для этих целей в данном случае используется гибридная (объединённая) слуховая модель, включающая следующие части: модель наружного и среднего уха, кохлеарную модель, описывающую функционирование улитки уха человека и модель слухового нерва – рис. 10.2.

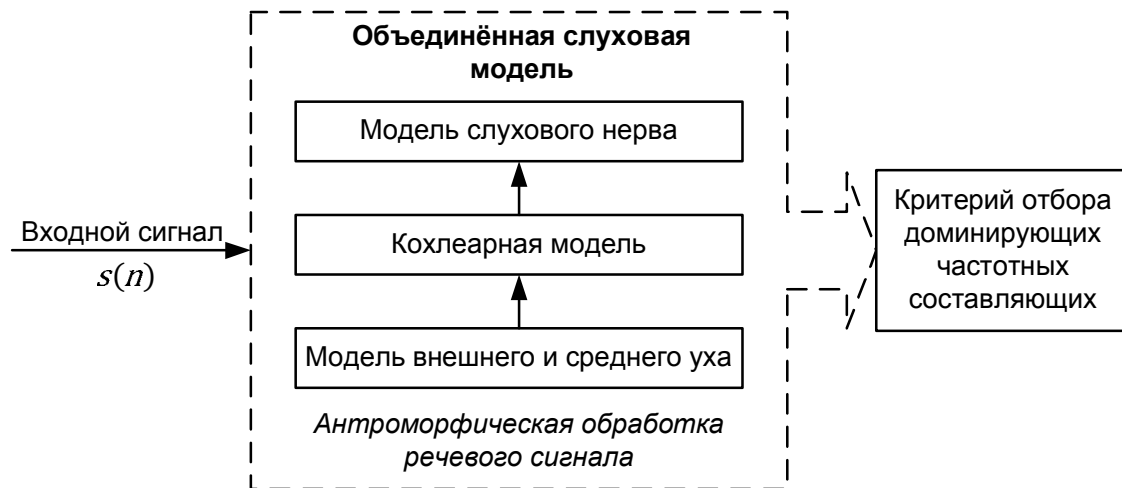


Рис. 10.2 – Обобщённая схема выделения доминирующих частотных составляющих на основе объединённой слуховой модели

Задача моделирования слуховой системы человека сводится к математическому описанию процессов, происходящих в различных частях слухового аппарата человека на физиологическом уровне.

Наиболее простой моделью для имитации эффектов изменения формы спектра сигнала, проходящего через внешнее и среднее ухо, является набор фильтров [12]: фильтр направ-

ления и фильтр предобработки. Амплитудно-частотная характеристика совмещённого фильтра наружного и среднего уха изображена на рис. 10.3.

При моделировании функционирования улитки уха человека необходимо математически описать процесс движения базилярной мембраны под воздействием акустических колебаний и тем или иным способом выполнить спектральное разложение, которое является основной функцией примыкающих к базилярной мембране внутренних волосковых клеток (ВВК) [13-15] – рис. 10.4.

Для моделирования работы слухового нерва в данном случае достаточно описать функционирование внутренних волосковых клеток и симитировать величину их нервного возбуждения на различных участках базилярной мембраны [16,17]. По степени возбуждения можно будет судить о том, насколько важна та или иная составляющая сигнала для человеческого слуха.

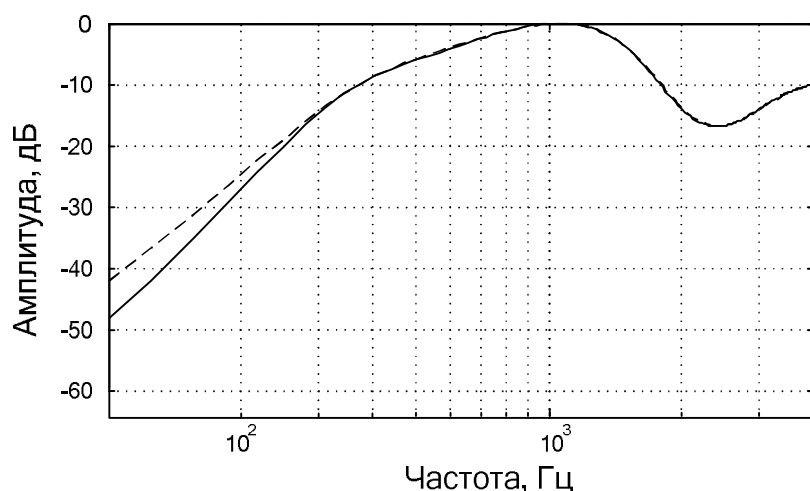


Рис. 10.3 – Амплитудно-частотная характеристика фильтра наружного и среднего уха: сплошная линия – экспериментальные данные из [12]; непрерывная линия – аппроксимация с помощью БИХ-фильтра 20-го порядка

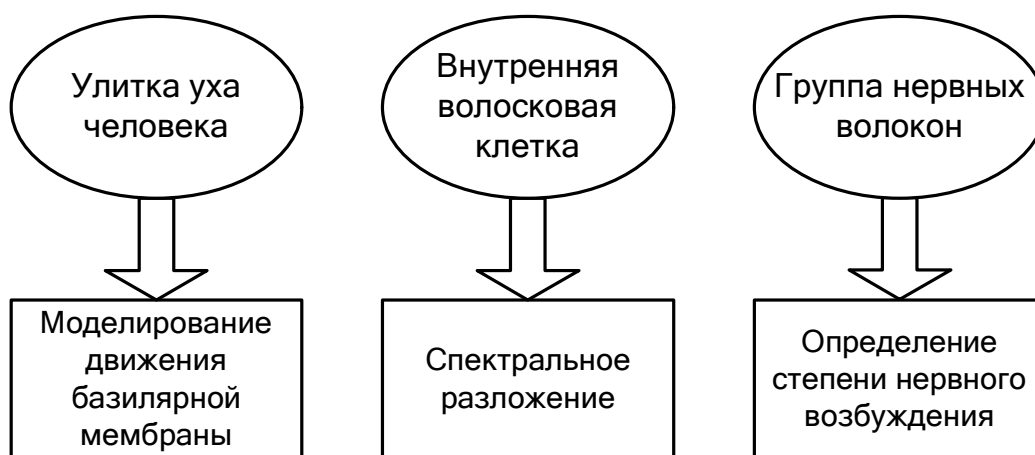


Рис. 10.4 – Обобщённая модель слухового аппарата человека

Таким образом, типичную структуру слуховых моделей [2,18,19], которые обладают высокой степенью адекватности физиологическим процессам можно представить, как показано на рис. 10.5.

В данном случае предполагается, что обработка сигнала строится преимущественно в частотной области. Это позволяет эффективно использовать полученные результаты совместно с синусоидальной моделью.

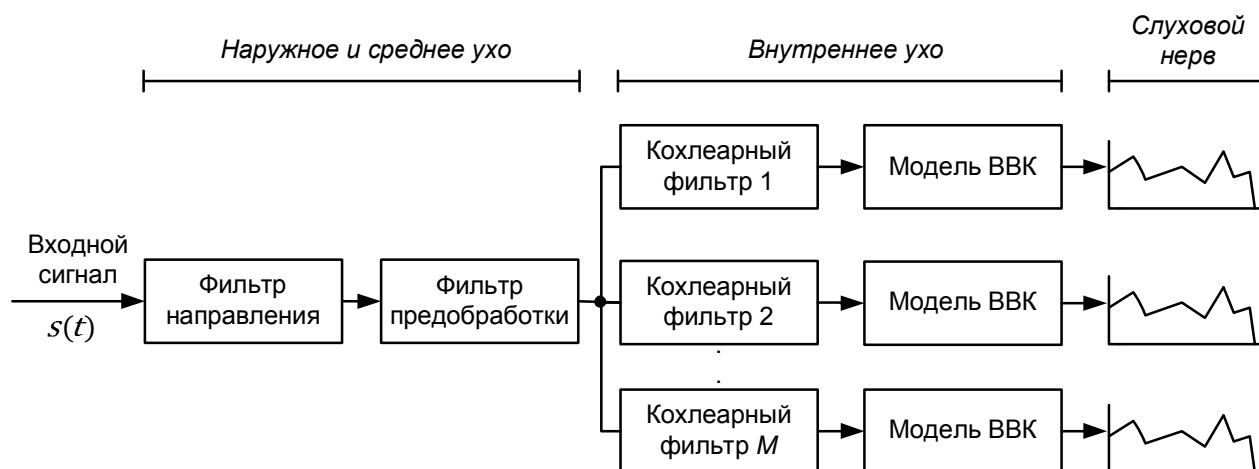


Рис. 10.5 – Типичная структура слуховой модели

10.2.2 Пассивная кохлеарная модель

В качестве основной модели периферической части слуховой системы человека используется SDCM-модель – Second Order Difference Cochlea Model (разностная кохлеарная модель второго порядка) [20].

Данная модель представляет собой двумерную математическую модель улитки человеческого уха, основанной на описании параметров движения вязкой несжимаемой жидкости. Она описывает процесс движения базилярной мембраны под воздействием акустических колебаний и объясняет механизм спектрального разложения сигнала. Согласно данной модели функционирование улитки уха описывается следующим разностным уравнением:

$$y_m(n) + b_{1m}y_m(n - 1) + b_{2m}y_m(n - 2) = A_m a_{0m}[u_s(n) - u_s(n - 2)], \quad (10.1)$$

где $y_m(n)$ – перемещение или так называемая пучность базилярной мембраны в позиции x_m ; b_{1m} , b_{2m} , A_m и a_{0m} – параметры, которые определяются физическими свойствами базилярной мембраны в позиции x_m и изменяются вдоль базилярной мембраны; m – номер сегмента базилярной мембраны после дискретизации; $u_s(n)$ – входной синусоидальный сигнал, характеризующий скорость перемещения стремечка.

Передаточную функцию модели улитки в дискретном пространстве и времени $H_m(z)$ можно записать в виде:

$$H_m(z) = \left[A_m \frac{a_{0m}(1 - z^{-2})}{1 + b_{1m}z^{-1} + b_{2m}z^{-2}} \right]^{NS}, \quad (10.2)$$

где NS – количество секций фильтра (в соответствии с данной моделью $NS = 2$ [20]).

Для примера на рис. 10.6 отображены амплитудно-частотные характеристики для 32 кохлеарных фильтров, а характеристика полосы пропускания по уровню 3 дБ представлена на рис. 10.7.

На рис. 10.7 горизонтальная координата определяет изменение центральной частоты кохлеарного фильтра, а вертикальная координата – соответствующую ей полосу пропускания кохлеарного фильтра. В данном случае при вычислении амплитудно-частотных характеристик фильтров минимально возможная полоса пропускания принималась равной 80 Гц. На рис. 10.8 изображена кохлеарная карта модели – отношение между частотой возбуждения (центральной частотой) и тем положением на базилярной мембране, которое претерпевает наибольшее смещение.

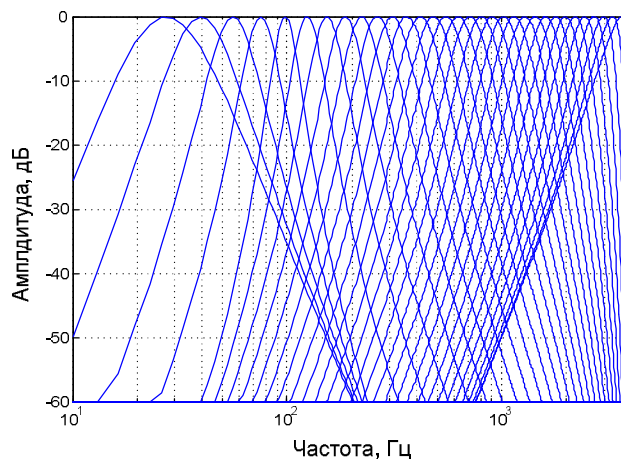


Рис. 10.6 – Амплитудно-частотные характеристики для 32 кохлеарных фильтров

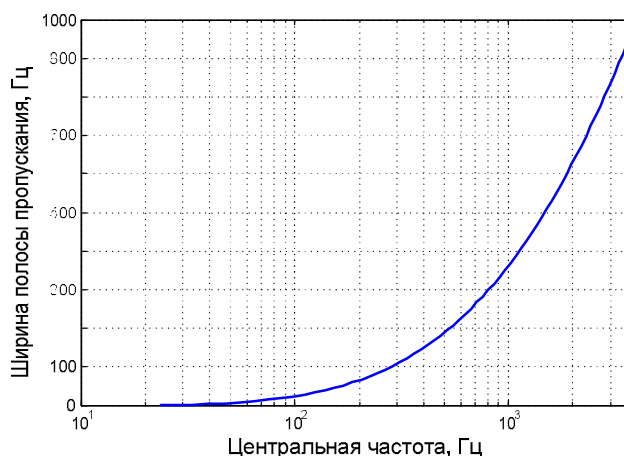


Рис. 10.7 – Характеристика полосы пропускания по уровню 3 дБ

Как видно из рис. 10.8 характеристики данной модели хорошо согласуются с экспериментальными данными [21,22].

Для повышения степени адекватности реальному процессу распространения акустических колебаний в периферической системе слуха человека необходимо произвести коррекцию описанной кохлеарной модели с учётом передаточной характеристики внешнего и среднего уха.

Хорошо известно, что каскадное (последовательное) включение фильтров с передаточными характеристиками $H_1(z)$ и $H_2(z)$ даёт эквивалент фильтра с передаточной характеристикой $H_3(z)$ [23]:

$$H_3(z) = H_1(z) \cdot H_2(z) . \quad (10.3)$$

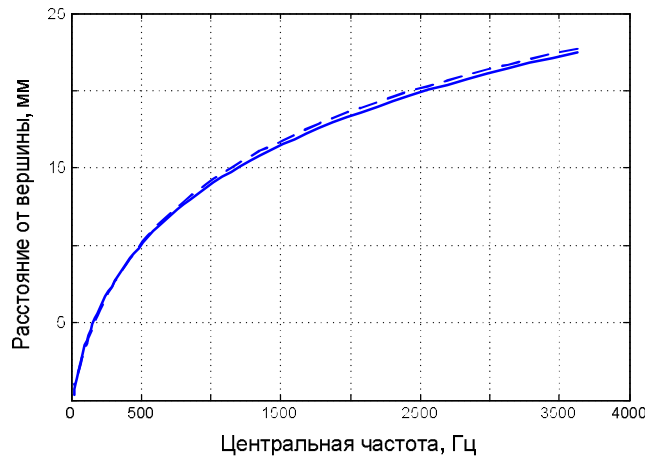


Рис. 10.8 – Кохлеарная карта: штриховая линия – экспериментальные данные из [21]; непрерывная линия – используемая модель

Тогда, при совмещении фильтра наружного и среднего уха и кохлеарных фильтров как показано на рис. 10.9 получим:

$$H_m^E(z) = H_{E1}(z) \cdot H_m(z), \quad m = \overline{1, M_F}, \quad (10.4)$$

где $H_m^E(z)$ – передаточная функция эквивалентного фильтра; $H_{E1}(z)$ – передаточная функция фильтра наружного и среднего уха; $H_m(z)$ – передаточная функция m -го кохлеарного фильтра.

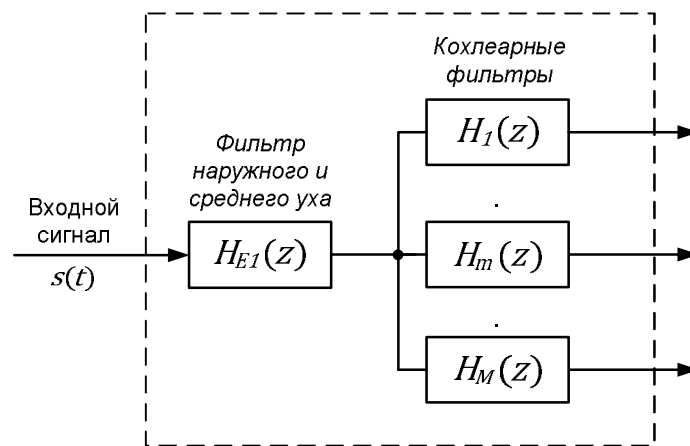


Рис. 10.9 – Схема включения фильтра наружного и среднего уха и кохлеарных фильтров

На основании выражения (10.2) можно записать:

$$H_m^E(z) = H_{E1}(z) \left[A_m \frac{a_{0m}(1 - z^{-2})}{1 + b_{1m}z^{-1} + b_{2m}z^{-2}} \right]^{NS}. \quad (10.5)$$

Пример амплитудно-частотных характеристик для 32 модифицированных кохлеарных фильтров отображены на рис. 10.10. При вычислении данных амплитудно-частотных характеристик фильтров величина NS принималась равной 2.

При организации обработки сигнала в частотной области необходимо учитывать ещё один важный момент. В настоящее время известно, что свойства восприятия акустических сигналов человеком на различных частотах неодинаковы. Измерения характеристик возбуждения волокон нерва в слуховом аппарате млекопитающих показывает значительное различие между

свойствами слуха на низких (меньше 1000 Гц) и высоких частотах [18,24,25]. На низких частотах гармонические компоненты сигнала распознаются более точно, т.е. на высоких частотах разрешение хуже. Считается, что эти различия в основном обусловлены механическими свойствами базилярной мембраны.

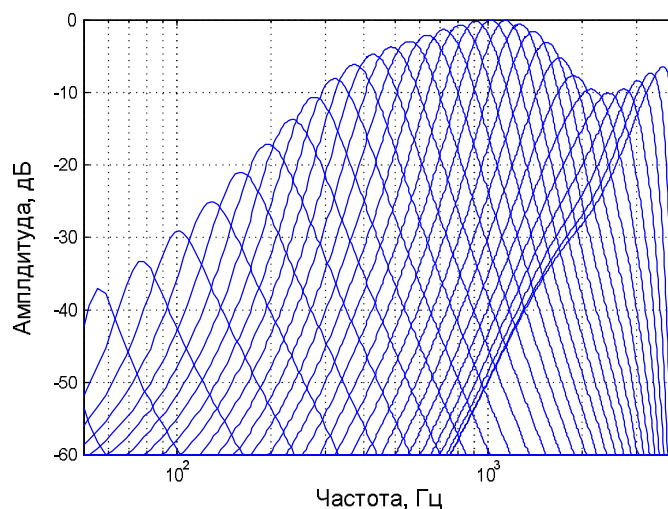


Рис. 10.10 – Амплитудно-частотные характеристики для 32 модифицированных кохлеарных фильтров

Поэтому для повышения степени адекватности процедуры антропоморфической обработки в частотной области используется ДПФ с неравномерным частотным разрешением (с неравным интервалом дискретизации по оси частот) [26]. Это позволяет использовать для анализа частотную шкалу, которая соответствует особенностям восприятия человека – шкалу барков [10]. Данный подход не только достаточно хорошо учитывает особенности человеческого слуха, но и делает возможным корректный анализ сигнала на частотах ниже 100 Гц.

10.2.3 Активная кохлеарная модель

Представленная модель улитки уха человека хорошо согласуется с экспериментальными данными, но не вполне приемлема для анализа сигналов, имеющих большой динамический диапазон (десятки дБ). Это объясняется следующими физиологическими особенностями [9].

Слуховая система человека характеризуется довольно широким рабочим диапазоном [10]: от 0 до 100-120 дБ. При этом сигнал с уровнем 60 дБ порождает в улитке уха колебания с очень малой амплитудой – всего несколько нанометров, что примерно соответствует величине диаметра атома. При такой малой величине смещения базилярной мембраны очень сложно объяснить механизм возникновения раздражающего воздействия на волосковые клетки, которые порождают электрические потенциалы, превышающие уровень их спонтанной активности в состоянии покоя. Кроме того, в работе [27] было показано, что отклик базилярной мембраны при возбуждении слабыми звуками намного более точно сконцентрирован на участке с соответствующей характеристической частотой, чем при воздействии сигналами высокой интенсивности. Получается, что полоса пропускания кохлеарного фильтра, который эквивалентен данному участку базилярной мембраны, меняется для сигналов различной силы. Наименьшая по-

лоса пропускания наблюдается при воздействии сигналов с небольшим уровнем и расширяется по мере возрастания уровня сигнала. В работах [28,29] было экспериментально показано, что в улитке уха млекопитающих существует некий активный процесс (кохлеарное усиление), который и объясняет вышеприведенные особенности функционирования органа слуха. Попытки смоделировать данный процесс были сделаны в работах [30,31], где была предложена активная модель улитки. В работе [32] предлагается рассматривать активную кохлеарную модель как пассивную модель улитки с обратной связью, реализуемой с помощью внешних волосковых клеток – рис. 10.11.

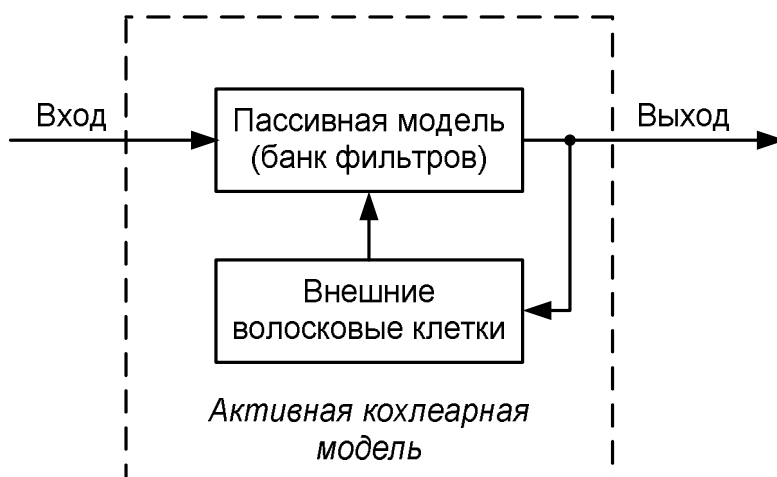


Рис. 10.11 – Схема активной кохлеарной модели

В настоящее время достаточно широко используются активные кохлеарные модели, предложенные в ряде работ [33-36], в которых модель улитки представлена в виде совокупности банка фильтров с постоянными коэффициентами, имитирующий отклик базилярной мембраны на возбуждение определённой фиксированной величины, с банком перестраиваемых фильтров. Коэффициенты перестраиваемых фильтров изменяются с помощью механизма обратной связи в виде модели внешних волосковых клеток – рис. 10.12.

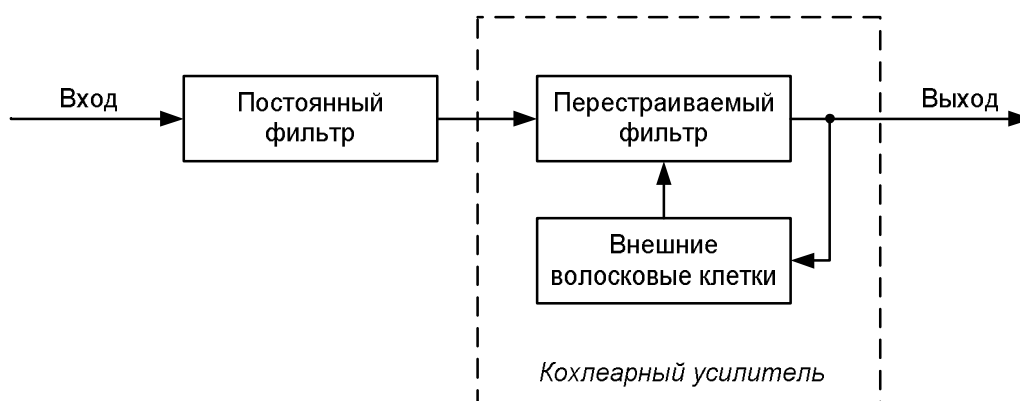


Рис. 10.12 – Схема активной кохлеарной модели

Следовательно, основной задачей модели внешних волосковых клеток является регулировка коэффициентов усиления перестраиваемых фильтров в банке таким образом, чтобы обеспечивать максимальное усиление при поступлении на вход сигнала наименьшей интен-

сивности и уменьшать значения коэффициентов при возрастании мощности частотных составляющих входного сигнала.

Основным достоинством применяемых в настоящее время подходов для проектирования активных кохлеарных моделей является довольно высокая степень адекватности физиологическим процессам, однако они обладают относительно высокой вычислительной сложностью.

Кроме того, при использовании этих моделей во временной области также требуются специальные механизмы для учёта фазовых параметров, что часто приводит к снижению степени их адекватности. С другой стороны, при компрессии речевого сигнала можно ограничиться определением усреднённых речевых параметров за определённый промежуток времени и отслеживать изменение этих параметров через фиксированные интервалы времени. Поэтому в данном случае предлагается модифицировать процесс обработки речевого сигнала с помощью активной кохлеарной модели следующим образом.

Как известно любой цифровой фильтр представляет собой работающую в дискретном времени систему с постоянными параметрами, причём сигналы на входе $x(n)$ и выходе $y(n)$ связаны дискретной свёрткой:

$$y(n) = x(n) * h(n), \quad (10.6)$$

где $h(n)$ – единичный импульс, а символ $*$ обозначает свёртку.

Соответствующее соотношение между z -преобразованиями входа $X(z)$ и выхода $Y(z)$ имеет следующий вид:

$$Y(z) = X(z) \cdot H(z), \quad (10.7)$$

где $H(z)$ – передаточная функция фильтра.

Учитывая, что дискретизированным вариантом z -преобразования является дискретное преобразование Фурье можно записать:

$$Y(k) = X(k) \cdot H(k), \quad (10.8)$$

где k – номер частотного отсчёта.

Поскольку результат фильтрации $Y(k)$ в данном случае используется только для анализа амплитуды на фиксированных частотах, то фазовую составляющую можно не учитывать:

$$|Y(k)| = |X(k)| \cdot |H(k)|. \quad (10.9)$$

Тогда с учётом выражения (10.9) результат работы m -го канала пассивной части кохлеарной модели $A_m(k)$ может быть представлен следующим выражением:

$$A_m(k) = |S(k)| \cdot H_m^E(k), \quad (10.10)$$

где $H_m^E(k)$ – АЧХ m -го модифицированного кохлеарного фильтра; $|S(k)|$ – спектр речевого сигнала.

Процесс работы кохлеарного усилителя в частотной области можно симитировать домножением результата фильтрации $A_m(k)$ на взвешивающий коэффициент $\alpha_m(A)$, зависящий в свою очередь от величины произведения $|S(k)| \cdot H_m^E(k)$ – рис. 10.13.

Из работ [2,13] известно, что частотная характеристика кохлеарного фильтра зависит от интенсивности сигнала на его входе. Для амплитуды входного сигнала, равной 0 дБ усиление фильтра на центральной частоте приблизительно равно 40 дБ. При увеличении интенсивности входного сигнала постепенно увеличивается полоса пропускания фильтра и незначительно “смещается” центральная частота, а усиление уменьшается и стремится к нулю при амплитуде сигнала свыше 100 дБ. Такое поведение кохлеарного фильтра соответствует физиологическим особенностям восприятия звука человеком [13].

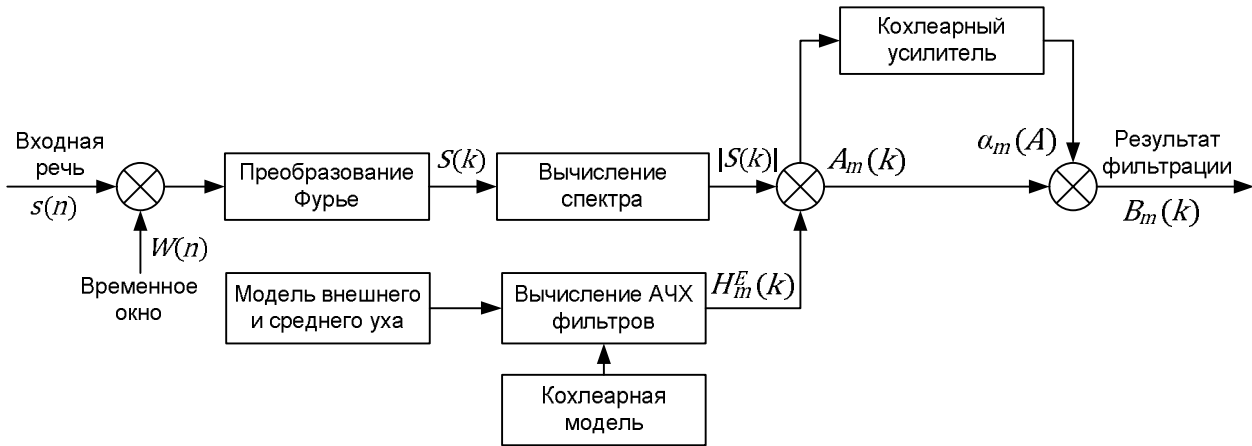


Рис. 10.13 – Схема активной кохлеарной модели в частотной области

Однако в устройствах обработки речевых сигналов нет необходимости настолько точно учитывать изменение полосы пропускания фильтра и смещение его центральной частоты, поскольку эти эффекты сильно выражены при уровне входного сигнала свыше 80 дБ, а для кодирования речи с хорошим качеством достаточно динамического диапазона до 60 дБ [37].

Поэтому в данном случае работу кохлеарного усилителя при его использовании для анализа речевого сигнала можно описать следующим выражением:

$$\alpha_m(A) = 40 \cdot (1 - A), \quad (10.11)$$

где $\alpha_m(A)$ – коэффициент усиления для m -го фильтра в дБ; A – амплитуда результата фильтрации $A_m(k)$ в дБ.

10.2.4 Модель слухового нерва

Очень хорошей степенью адекватности реальным физиологическим процессам обладает модель слуха человека, представленная в работе [2] и основанная на вычислении групповой интервальной гистограммы (Ensemble Interval Histogram) – $G(t, f)$ (рис. 10.14).

В данной модели группа нервных волокон, раздражаемых в одной внутренней волосковой клетке, имитируется массивом детекторов пересечения уровня $1 \dots P$, расположенных на выходе каждого кохлеарного фильтра [2]. Каждый детектор пересечения уровня является своеобразным эквивалентом нервного волокна в волосковой клетке. Силу нервного возбуждения на физиологическом уровне имитирует определенное пороговое значение уровня пересечения для каждого детектора. Результат процесса пересечения уровня может быть интерпретирован как совместное нервное возбуждение группы нервных волокон в разных внутренних волосковых клетках, расположенных вдоль базилярной мембраны в одной и той же области (сегменте

базиллярной мембраны), свойство которой отражает частотная характеристика одного кохлеарного фильтра.

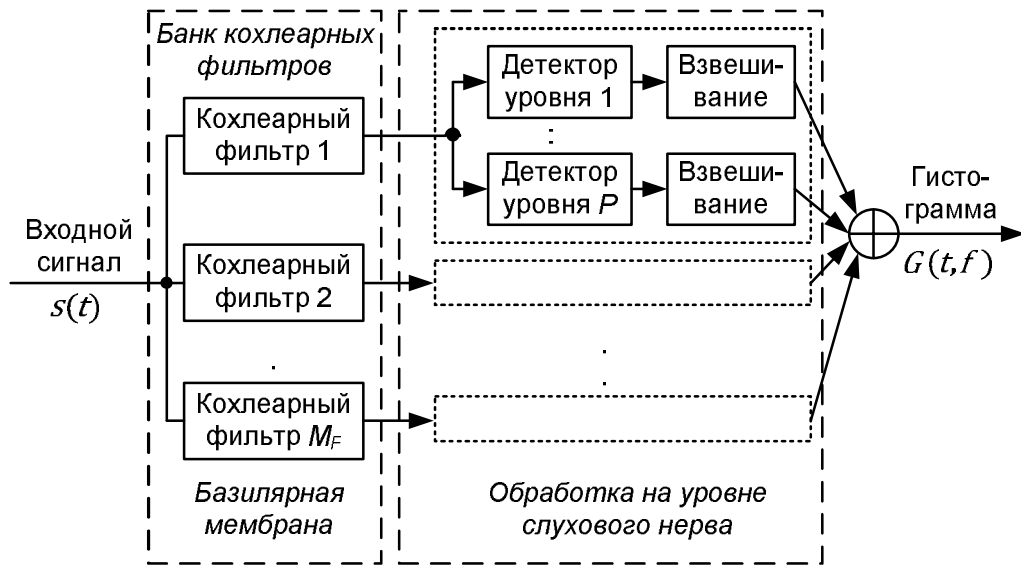


Рис. 10.14 – Модель слухового нерва во временной области

Данная модель системы слуха человека позволяет получить представление об акустической информации, циркулирующей на уровне слухового нерва. С её помощью можно дифференцировать частотные составляющие анализируемого речевого сигнала по степени их “важности” для человеческого слуха. Кроме того, она учитывает некоторые эффекты маскирования [10,16]. Однако, непосредственное применение этой модели в данном случае затруднительно из-за её плохой “совместимости” с синусоидальной моделью. Кроме того, она обладает достаточно высокой вычислительной сложностью, что также усложняет обработку речевого сигнала на её основе в реальном масштабе времени.

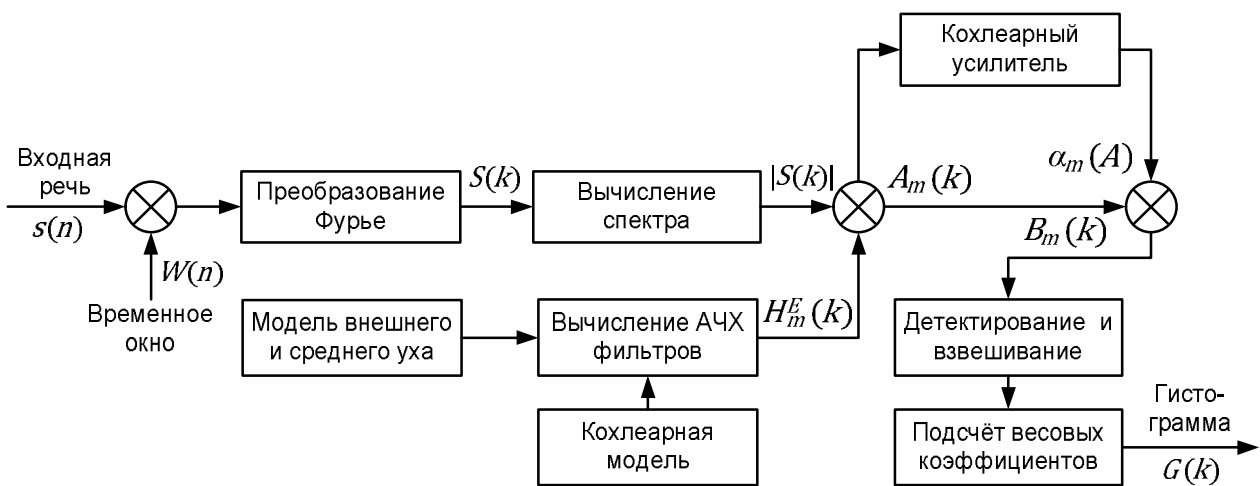


Рис. 10.15 – Схема вычисления модифицированной слуховой гистограммы

Поэтому, в данном случае процесс вычисления гистограммы $G(t, f)$ модифицируется таким образом, чтобы устранить вышеперечисленные недостатки и в тоже время сохранить её “полезные” свойства. Одним из возможных решений является перенос большинства действий из временной области в частотную, что позволяет не только резко снизить вычислительную

сложность алгоритма анализа, но и даёт возможность корректно совместить его с применяемым для определения синусоидальных параметров спектральным анализом.

Применяя обработку входного речевого сигнала с помощью активной кохлеарной модели в частотной области интервальную гистограмму $G(t, f)$ можно представить в виде дискретной функции частоты $G(k)$ – рис. 10.15.

В данном случае гистограмма $G(k)$ вычисляется с помощью следующего выражения:

$$G(k) = \sum_{m=1}^{M_F} G_m(k), \quad k = 1, \frac{\overline{N_F}}{2} \quad (10.12)$$

где m – номер обрабатываемого в текущий момент времени кохлеарного канала; M_F – число кохлеарных фильтров; N_F – размер ДПФ; $G_m(k)$ – k -й элемент гистограммы для m -го кохлеарного фильтра, он может быть вычислен по следующей формуле:

$$G_m(k) = |S(k)| \cdot H_m^E(k) \cdot \alpha_m(A) \sum_{l=1}^L p_m(l), \quad (10.13)$$

где L – количество уровней; $\alpha_m(A)$ – коэффициент усиления для m -го кохлеарного фильтра; $p_m(l)$ – весовая функция, которая вычисляется по следующему условию:

$$p_m(l) = \begin{cases} \beta_l, & \text{если } |S(k)| \cdot H_m^E(k) \cdot \alpha_m(A) \geq \text{Amp}_l, \\ 0, & \text{если } |S(k)| \cdot H_m^E(k) \cdot \alpha_m(A) < \text{Amp}_l, \end{cases} \quad (10.14)$$

где Amp_l – амплитуда l -го уровня; β_l – постоянная величина, характеризующая степень нервного возбуждения, каждому уровню сопоставляется своё собственное значение β_l .

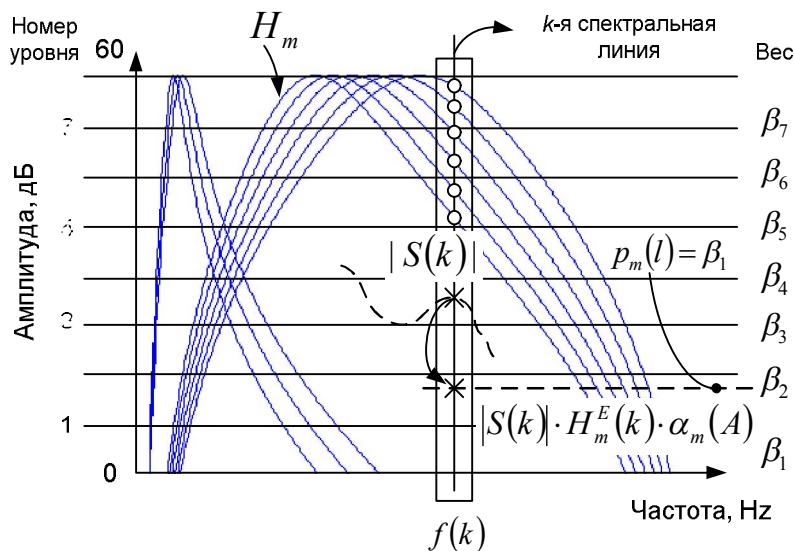


Рис. 10.16 – Иллюстрация процесса вычисления k -го элемента гистограммы в частотной области

На основании (10.11) работу кохлеарного усилителя можно представить следующим образом:

$$\alpha_m(A) = \frac{100}{10^{40 \log_{10}[|S(k)| \cdot H_m^E(k)]}} \quad (10.15)$$

Пример вычисления одного элемента гистограммы показан на рис. 10.16. Из рис. 10.16 видно, что вес выбирается в зависимости от величины произведения $|S(k)| \cdot H_m^E(k) \cdot \alpha_m(A)$. По аналогии с [2] в данном случае каждому уровню также сопоставляется фиксированное значение веса. Уровни равномерно распределены по всему амплитудному диапазону спектра сигнала.

10.3 Особенности построения системы компрессии речевого сигнала на основе синусоидальной модели с антропоморфической обработкой

10.3.1 Кодер

Структура синусоидального кодера речи с антропоморфической обработкой речевого сигнала представлена на рис. 10.17. Входной оцифрованный речевой сигнал анализируется с помощью спектрального анализа на основе преобразования Фурье, совмещённого с антропоморфической обработкой. В процессе анализа в исходном фрагменте речевого сигнала с помощью процедуры антропоморфического анализа выделяется фиксированное количество наиболее “важных” для человеческого слуха доминирующих синусоидальных компонент, для каждой из которых в дальнейшем определяется амплитуда, частота и фаза.

Процесс анализа речевого сигнала с использованием данного подхода можно представить следующим образом. Входной речевой сигнал $s(n)$ анализируется отрезками определённой длины. Фреймы сигнала берутся с некоторым наложением.

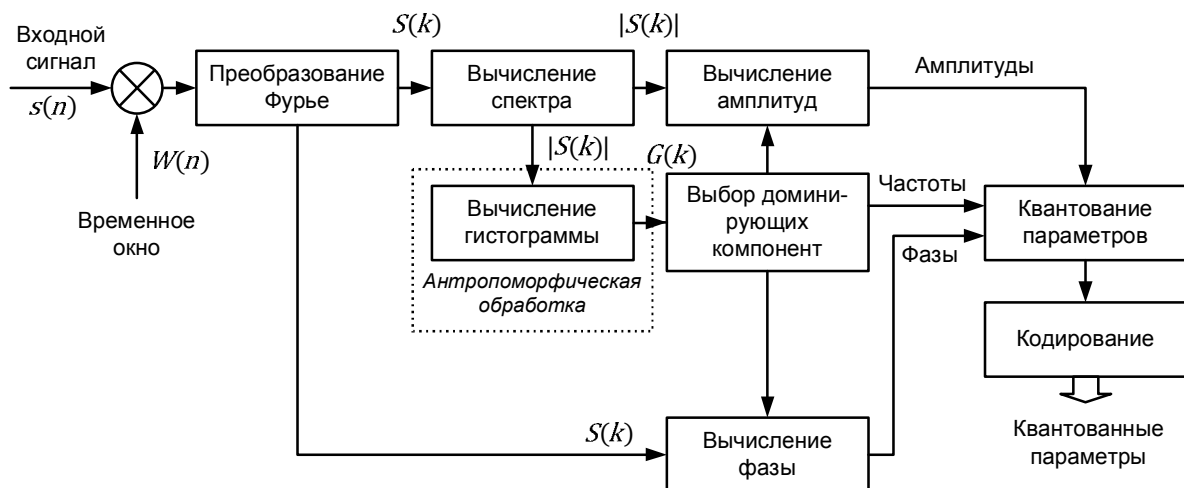


Рис. 10.17 – Структура кодера

Для каждого фрейма вычисляется модуль ДПФ – спектр сигнала $|S(k)|$. Отбор доминирующих синусоидальных составляющих осуществляется на основании слуховой гистограммы $G(k)$, с помощью которой все частотные составляющие анализируемого речевого сигнала дифференцируются по степени их важности для человеческого слуха. Каждый элемент гистограм-

мы $G(k)$ представляет собой значение веса, который имитирует силу нервного возбуждения для k -ой частотной составляющей. Частоты синусоид определяются по спектру сигнала $|S(k)|$ исходя из местоположения выбранных частотных составляющих на оси частот, а по их величине – амплитуды. Фазы определяются по действительной и мнимой компонентам полученного спектра на соответствующих частотах. Найденные таким образом параметры отобранных синусоидальных составляющих используются для представления речи на текущем фрейме.

Для передачи по линии связи полученные синусоидальные параметры должны быть заквантованы, т.е. должны быть представлены минимальным количеством бит, необходимым для сохранения хорошего качества синтезированного сигнала. Для вышеописанной системы фазы имеют равномерное распределение, а амплитуды и частоты – нет [38]. Кроме того, наблюдается значительное отличие в распределении амплитуд и частот для разных составляющих из набора, приходящихся на один фрейм.

Поскольку частоты и амплитуды определяются по спектральной характеристике речевого сигнала, для их кодирования целесообразно использовать векторное квантование в пространстве амплитуда-частота. На сторону декодера в качестве параметра, вместо непосредственных значений амплитуды и частоты, передаётся только индекс найденного элемента в кодовой книге. Фазы кодируются с использованием скалярного квантования. Амплитуды кодируются в логарифмическом диапазоне от 0 до 60 дБ с равномерным шагом. Этого вполне достаточно для кодирования речевого сигнала с хорошим качеством. Поскольку применяется двумерное векторное квантование, то производится согласование шага квантования по амплитуде и частоте [38].

Процесс квантования амплитуды и частоты для одной синусоидальной составляющей может быть описан следующим образом. Из определённых в процессе анализа значений амплитуды A и частоты k для каждой синусоидальной компоненты формируется входной вектор. В заранее сформированной кодовой книге C с набором векторов y_i осуществляется поиск, используя критерий минимальной ошибки квантования между входным вектором и вектором из кодовой книги. Когда такой вектор найден, его индекс кодируется и передаётся на сторону декодера, который содержит копию кодовой книги из кодера. Кодовая книга представляет собой набор из L предопределённых выходных векторов $y_i = \{A_i, k_i\}$. Для поиска оптимального элемента в кодовой книге используется критерий ближайшего элемента (nearest neighbor condition). Для тренировки кодовой книги используется итерационный кластерный алгоритм, широко известный как алгоритм K -средних (данный алгоритм известен также как обобщенный алгоритм Ллойда – Generalized Lloyd Algorithm) [39].



Рис. 10.18 – Схема коррекции частоты

Описанный выше подход для квантования синусоидальных параметров даёт хорошие результаты, если длина кодовой книги равняется 4096 или больше [38]. Однако в этом случае данный алгоритм квантования параметров имеет высокую вычислительную сложность, что создаёт трудности для его использования в системах реального времени. Также, если длина кодовой книги очень велика, то возникают трудности и в процессе тренировки кодовой книги (используется большое тренируемое множество и т.п.). Если кодовая книга слишком мала, то выходная синтезированная речь имеет плохое качество, появляются значительные искажения. Экспериментальным образом установлено, что в данной системе именно ошибка по частоте чрезвычайно сильно влияет на качество синтезируемой речи [38]. Поэтому предлагается комбинировать векторное квантование амплитуд и частот со скалярным квантованием ошибки по частоте. Коррекцию частоты в этом случае можно представить как показано на рис. 10.18.

Фазы квантуются используя скалярное квантование. Учитывая периодичность тригонометрических функций, фазы могут квантоваться только в диапазоне от $-\pi$ до π . Для получения хороших результатов при таком подходе для представления фаз достаточно 2-3 бит на одну составляющую [38].

Приемлемые результаты получаются при следующих характеристиках: длина кодовой книги для квантования амплитуды и частоты – 256 элементов (8 бит), для кодирования ошибки квантования частоты используется 4 бита и для квантования фазы – 3 бита. Таким образом, для представления параметров одной синусоидальной компоненты требуется 15 бит. Если анализ-синтез речевого сигнала происходит с частотой 50 раз в секунду при длине фрейма синтезируемой речи 20 мс и количестве синусоидальных компонент от 5 до 10 штук использование описанного метода квантования параметров позволяет достичь скорости передачи данных по каналу от 3750 до 7500 бит/с.

10.3.2 Декодер

На стороне декодера принятые параметры деквантуются и декодируются, а сам процесс синтеза речи сводится к суммированию сгенерированных синусоидальных компонент с найденными в процессе анализа амплитудами, фазами и частотами. Однако, поскольку параметры речевого сигнала изменяются во времени, а сам сигнал представляется на отдельно взятых фреймах ограниченным (не бесконечным) набором синусоид с постоянными частотами, амплитудами и фазами, то на границах синтезируемых фреймов генерируемый речевой сигнал будет меняться скачкообразно. Это приводит к возникновению хриплости и появлению различных искажений в синтезированной речи. Следовательно, возникает задача “склеивания” соседних фреймов и устранения разрывов на их границах.

Для устранения этих нежелательных явлений в данном случае применяется метод наложения со сложением (overlap-add), при котором каждый генерируемый фрагмент речевого сигнала определённым образом накладывается на предыдущий. Процесс синтеза при использовании метода наложения со сложением можно представить как показано на рис. 10.19.

Главное достоинство данного метода синтеза речевого сигнала состоит в простоте реализации, он имеет низкую алгоритмическую и вычислительную сложность. Данный подход также применим и в ситуациях, когда требуется проводить анализ-синтез сигнала с малой дли-

ной фрейма (менее 20 мс), даже, если синусоидальные параметры при этом удовлетворяют условию строгой стационарности.

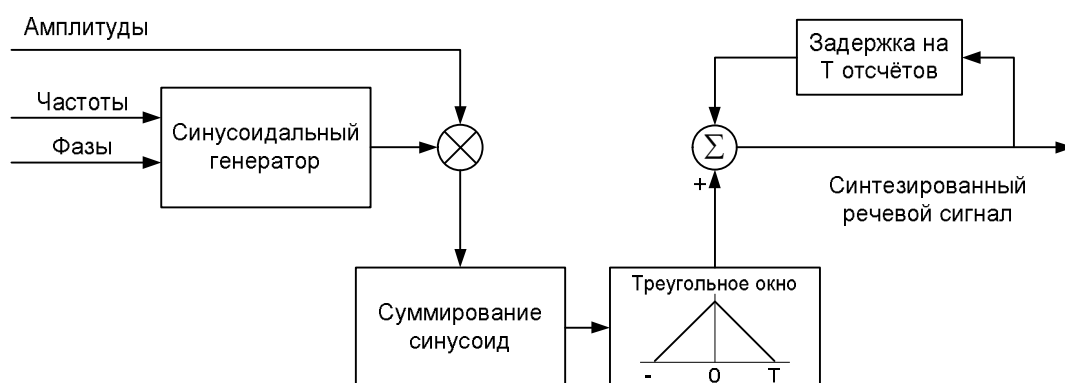


Рис. 10.19 – Схема синусоидального синтеза речевого сигнала при использовании метода наложения со сложением

Для примера на рис. 10.20 отображён отрезок данного речевого сигнала во временной области с нормированной амплитудой от -1 до 1 и длительностью около 1200 мс, которому соответствует слово на русском языке “недорогой”. Частота дискретизации сигнала $F_s = 8000$ Гц, амплитуда – 16 бит на отсчёт, количество каналов – один, голос – мужской. На рис. 10.21 отображена спектрограмма данного отрезка речевого сигнала. По оси абсцисс отложено время в секундах (временной масштаб согласован с сигналом на рис. 10.20), а по оси ординат – частота в диапазоне от 0 до половины частоты дискретизации $F_s/2$ Гц. На рис. 10.22 изображён отрезок синтезированного речевого сигнала. При синтезе использовались следующие параметры: количество синусоидальных составляющих – 10; длина преобразования Фурье – 1024; длительность окна анализа – 32 мс; длительность фрейма синтеза – 20 мс. На рис. 10.23 изображена спектрограмма данного отрезка.

С помощью специального программного обеспечения [40] были получены объективные оценки качества восстановленного речевого сигнала для описываемой системы компрессии:

- 1) отношение сигнал/шум (SNR): 8,36 дБ;
- 2) отношение шум/порог маскирования (NMR): -3,47 дБ;
- 3) искажение спектра барков, вычисленное по стандартному алгоритму (BSD): 0,09;
- 4) искажение спектра барков, вычисленное по модифицированному алгоритму ($MBSD$): 0,07.

Для данного теста в качестве речевых образцов использовались предварительно записанные и записанные в формате .WAV фразы. Были записаны голоса для пяти дикторов (три мужских и два женских).

Кроме того в соответствии с методикой, приведённой в [41], для синусоидального вокодера с антропоморфической обработкой при скорости передачи параметров 8000 бит/с были проведены следующие тесты:

- 1) измерение разборчивости речи (относительное количество правильно принятых элементов речи) артикуляционным методом;

2) измерение качества речи и узнаваемости диктора методом парных сравнений с контрольным трактом;

3) измерение степени узнаваемости голоса диктора (возможность слушателей отождествлять звучание голоса, принимаемого из испытуемого тракта, с конкретным лицом, известным слушателю ранее).

Для измерения разборчивости речи артикуляционным методом в качестве речевых образцов использовались предварительно надиктованные и записанные в формате .WAV наборы слогов из специальных артикуляционных таблиц [41]. Были записаны голоса для трёх дикторов (два мужских и один женский). Для прослушивания были задействованы три аудитора.

Для измерения качества речи и узнаваемости диктора методом парных сравнений с контрольным трактом и измерения степени узнаваемости голоса диктора были записаны и прослушаны голоса для пяти дикторов (три мужских и два женских). В качестве речевых образцов использовались предварительно надиктованные и записанные в формате .WAV следующие фразы [41].

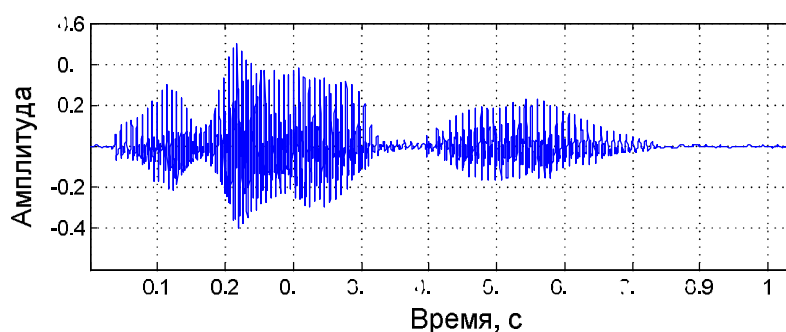


Рис. 10.20 – Отрезок оригинального речевого сигнала

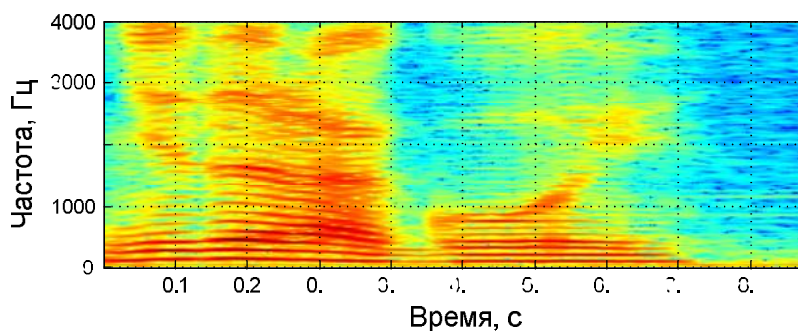


Рис. 10.21 – Отрезок синтезированного речевого сигнала

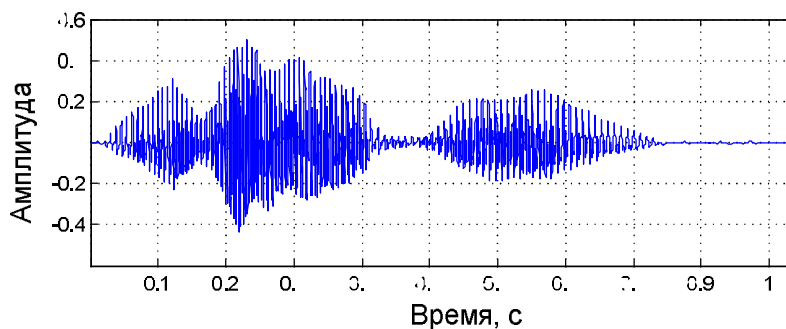


Рис. 10.22 – Спектрограмма отрезка оригинального речевого сигнала

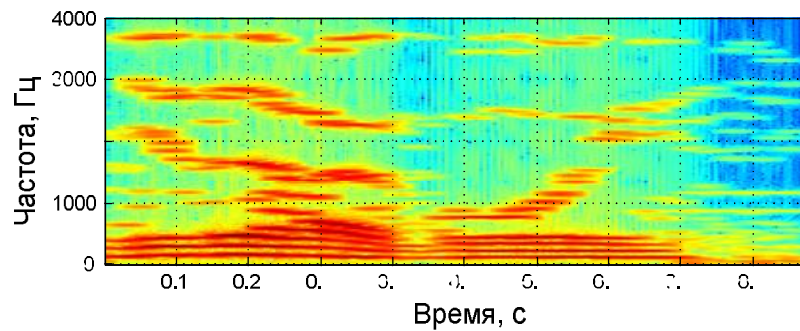


Рис. 10.23 – Спектрограмма отрезка синтезированного речевого сигнала

Получены следующие результаты:

- 1) слоговая разборчивость речи – 96 %;
- 2) субъективное качество речи – 3.8 баллов;
- 3) характеристика восстановленной в декодере речи – некоторое нарушение естественности и узнаваемости, иногда присутствуют подзванивание и дребезжание;
- 4) узнаваемость голоса диктора – 95 %.

Значения полученных объективных оценок являются типичными для низкоскоростных параметрических вокодерных устройств. Результаты субъективной оценки качества синтезированной речи подтверждают, что речь отличается довольно высокой степенью разборчивости и хорошей узнаваемостью диктора даже при ограниченном числе (до 10 штук) синусоидальных компонентов.

10.4 Список использованных источников

1. McAulay R.J., Quatieri T.F., Speech analysis/synthesis based on a sinusoidal representation, IEEE Trans. on Acoust., Speech and Signal Processing, 1986, Vol. ASSP-34, pp. 744–754.
2. Ghitza O., Auditory Models and Human Performance in Tasks Related to Speech Coding and Speech Recognition, IEEE Transactions on Speech and Audio Processing, 1994, Vol. 2, No. 1. pp. 115–132.
3. Ghitza O., Auditory Nerve Representation Criteria for Speech Analysis/Synthesis, IEEE Transactions on Speech and Audio Processing, 1987, Vol. ASSP-35, No. 6, pp. 736–740.
4. Wan W.G. and et al., A novel approach of low bit-rate speech coding based on sinusoidal representation and auditory model, Speech communication and technology: proceedings of 6-th europ. conf. (EUROSPEECH'99), Budapest, Hungary, September 5–9, 1999, pp. 1555–1558.
5. Edler B., Purnhagen H., Parametric Audio Coding, Signal processing: proceedings of 5-th int. conf. (ICSP'2000), Beijing, China, August 21–25, 2000, pp. 21–24.
6. Kondoz A.M., Digital speech: coding for low bit rate communication systems, New York: John Wiley & Sons Inc., 1996, 442 p.
7. Spanias A.S., Speech coding: a tutorial review, Proc. IEEE. 1994, Vol. 82, No. 10, pp. 1541–1582.

-
8. George E.B., Smith M.J.T., Perceptual considerations in a low bit rate sinusoidal vocoder, Computers and communications: proceedings of int. conf. (IPCCC'90), Phoenix, USA, March 21–23, 1990, pp. 268–275.
 9. Иванов А.В., Формирование пространства признаков на основе антропоморфической обработки информации в распознавателях речи в условиях противодействия: дис. канд. техн. наук: 05.13.17 / А.В. Иванов. – Минск, 2004. – 188 л.
 10. Zwicker E., Fastl H., Psychoacoustics: facts and models, Berlin: Springer-Verlag, 1990, 354 p.
 11. Verma T., Meng T., Sinusoidal modeling using frame-based perceptually weighted matching pursuits, Acoustics, speech and signal processing: proceedings of IEEE intern. conf. (ICASSP-99), Phoenix, Arizona, USA, March 15–19, 1999, Vol. 2, pp. 981–984.
 12. Glasberg B.R., Moore B.C.J., Derivation of auditory filter shapes from notched noise data, Hearing Research, August 1990, Vol. 47, pp. 103–138.
 13. Goldstein J.L., Modeling rapid waveform compression on the basilar membrane as a multiple-bandpass-nonlinearity filtering, Hearing Research, 1990, Vol. 49, pp. 39–60.
 14. Kates J.M., Accurate Tuning Curves in a Cochlear Model, IEEE Trans. on Speech and Audio Processing, 1993, Vol. 1, No. 4. pp. 453–462.
 15. Likhachov D.S., Petrovsky A.A., Improved auditory-based speech coding using psychoacoustic model based on a cochlear filter bank and an average localized synchrony detection, Computer information systems and industrial management applications / Eds.: K. Saeed, R. Mosdorf, Z. Sosnowski, Poland, Bialystok, 2003, pp. 11–19.
 16. Ghitza O., Auditory Nerve Representation as a Basis for Speech Processing, Advances in Speech Signal Processing / Eds. S. Furui, M.M. Sondhi, New York: Marcel Dekker, 1992, pp. 453–485.
 17. Ghitza O., Processing of spoken CVC's in the auditory: Psychophysics, J. Acoust. Soc. Am. 1993, Vol. 94, No. 5, pp. 2507–2516.
 18. Delgutte B., Kiang N.Y.S., Speech coding in the auditory nerve: Vowel-like sounds, J. Acoust. Soc. Am., 1984, Vol. 75, pp. 866–878.
 19. Seneff S., A joint synchrony/mean-rate model of auditory speech processing, Journal of Phonetics. Special Issue, 1988, Vol. 16(1), pp. 55–76.
 20. Petrovsky A.A. and et. al., An anthropomorphic speech processing based on the cochlear model and its application for coding task, International scientific journal of computing, 2004, Vol. 3, No. 1, pp. 75–83.
 21. Greenwood D.D., A cochlear frequency-position function for several species-29 years later, J. Acoust. Soc. Am., 1990, Vol. 87, No. 6, pp. 2592–2605.
 22. Rhode W.S., Robles L., Evidence from mossbauer experiments for nonlinear vibrations in the cochlea, J. Acoust. Soc. Am., 1974, Vol. 55, No. 3, pp. 588–596.

-
23. Гольденберг Л.М., Матюшкин Б.Д., Поляк М.Н., Цифровая обработка сигналов: Справочник, М.: Радио и связь, 1985, 312 с.
 24. Sachs M.B., Young E.D., Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate, *J. Acoust. Soc. Am.*, 1979, Vol. 66, pp. 470–479.
 25. Young E.D., Sachs M.B., Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers, *J. Acoust. Soc. Am.*, 1979, Vol. 66, pp. 1381–1403.
 26. Petrovsky A.A. and et al., Warped discrete Fourier transform in perceptual speech and audio processing, *New trends in audio and video: proceedings of X int. symposium, Wroclaw, Poland, 16–18 September 2004*, pp. 153–159.
 27. Rhode W.S., Some Observations on Cochlea Mechanics, *J. Acoust. Soc. Am.*, 1978, Vol. 64, No 1, pp. 158–176.
 28. Kemp D.T., Evidence of mechanical nonlinearity and frequency selective wave amplification in the cochlea, *Arch. Otorhinolaryngol*, 1979, Vol. 224, No. 1(2), pp. 37–45.
 29. Mountain D.C., Changes in endolymphatic potential and crossed olivocochlear bundle stimulation alter cochlear mechanics, *Science*, 1980, Vol. 210, No. 4465, pp. 71–72.
 30. Kim D.O.S., Neely S.T., An Active Cochlear Model with Negative Damping in the Cochlear Partition: Comparison with Rhode's ante- and post-mortem results, *Physiological and Behavioral Studies in Hearing / Eds. G. Brink, F.A. Bilsen, Delft: Delft University Pres, 1980*, pp. 7–14.
 31. Neely S.T., Kim D.O.S., An Active Cochlear Model Showing Sharp Tuning and High Sensitivity, *Hearing Research*, 1983, Vol. 9, No. 2, pp. 123–130.
 32. Mountain D.C. and et al., Electromechanical processes in the cochlea, *Mechanics of Hearing / Eds. E. Boer, M.A. Viergever, Delft: Delft University Press, 1983*, pp. 119–126.
 33. Eriksson J.L., Robert A., The Representation of Pure Tones and Noise in a Model of Cochlear Nucleus Neurons, *J. Acoust. Soc. Am.*, 1999, Vol. 106, No. 4, pp. 1865–1879.
 34. Kates J.M., Two-Tone Suppression in a Cochlear Model, *IEEE Transactions on Speech and Audio Processing*, 1995, Vol. 3, No. 5, pp. 396–406.
 35. Robert A., Eriksson J.L., A composite model of the auditory periphery for simulating responses to complex sounds, *J. Acoust. Soc. Am.*, 1999, Vol. 106, No. 4, pp. 1852–1864.
 36. Zhang X., Heinz M.G., A phenomenological model for the responses of auditory-nerve fibers: nonlinear tuning with compression and suppression, *J. Acoust. Soc. Am.*, 2001, Vol. 109, No. 2, pp. 648–670.
 37. Ivanov A.V. and et. al., Spiking neuron auditory model for speech processing systems, *Systems, signals and image processing: proceedings of the 9-th intern. workshop, Manchester, UK, November 7–8, 2002*, pp. 231–236.

-
38. Likhachov D.S., Petrovsky A.A., Parameters quantization in sinusoidal speech coder on basis of human auditory model, *Speech and Computer: proceedings of the 9-th intern. conf.*, Russia, Saint-Petersburg, September 20–22, 2004, pp. 195–202.
 39. Makhoul J. and et al., Vector quantization in speech coding, *Proc. IEEE.*, November 1985, Vol. 73, pp. 1551–1588.
 40. Петровский Ал.А., Объективная оценка качества восстановленного аудиосигнала перцептуальным ПДВП-кодером на базе периферийной модели уха человека, Сборник трудов 5-й Международной научной конференции “Цифровая обработка сигналов и её применение” (DSPA’2003), т.2, Москва, Россия, 2003. с. 123-126.
 41. Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости: СТБ ГОСТ Р 50840-2000, Введ. 01.04.2004, Минск: Белстандарт, 2000, 372 с.

11 ШИРОКОПОЛОСНЫЙ CELP-КОДЕР С МУЛЬТИПОЛОСНЫМ ВОЗБУЖДЕНИЕМ ПО БАНКУ КОДОВЫХ КНИГ С РЕКОНФИГУРИРУЕМОЙ АРХИТЕКТУРОЙ

Лившиц М.З., Парфенюк М., Петровский А.А.

АННОТАЦИЯ

В главе описывается модель и структура широкополосного перцептуального CELP-кодера с мультиполосным возбуждением и обработкой субполос в порядке убывания их перцептуальной важности. Для управления процессом кодирования используется психоакустическая модель на базе Дискретного преобразования Фурье с неравномерным разложением коэффициентов по частотной оси (от англ. – Warped Discrete Fourier Transform – WDFT), осуществляется синтез алгоритма управления структурой банка субполосных кодовых книг. Описывается формирование, оптимизации и кодирование структуры банка кодовых книг векторов возбуждения, а также интеграция препроцессора редактирования шума в структуру компрессора. Дается краткая характеристика квантования параметром модели широкополосного перцептуального компрессора речевых сигналов.

Ключевые слова – мультиполосное возбуждение, психоакустика, CELP, WDFT.

СОДЕРЖАНИЕ

11.1	Модель широкополосного кодера.....	234
11.2	Психоакустическая модель для оценки перцептуальной значимости субполос	239
11.2.1	Формирование субполосных групп на основе анализа их перцептуальной значимости.....	239
11.2.2	Оценка порогов маскирования в WDFT-области.....	243
11.3	Математическая модель перцептуального CELP-кодера с реконфигурируемым банком кодовых книг	245
11.3.1	Синтез субполосных кодовых книг с варьируемым уровнем детализации (количеством векторов).....	250
11.4	Препроцессор повышения качества речи.....	252
11.5	Математическая модель комбинированной системы компрессора широкополосных речевых сигналов с препроцессором редактирования шума.....	254
11.6	Список использованных источников.....	258

11.1 Модель широкополосного кодера

Перцептуальный широкополосный CELP-кодер речевых сигналов должен иметь гибкую настраиваемую структуру, отражающую статистические и перцептуальные особенности кодируемого сигнала. Это позволит значительно устранить присутствующую избыточность в сигнале,

обеспечив тем самым максимальную степень сжатия с одновременным сохранением высокого качества звучания реконструированного сигнала. Несмотря на то, что предсказание частоты основного тона является очень эффективным для речи, оно не совсем применимо для музыки. Структура музыкального сигнала является очень сложной и не может быть смоделирована с применением одного фильтра-синтезатора основного тона (от англ. – Long Term Predictor – LTP). Поэтому, необходимо использовать LTP-фильтр только тогда, когда это необходимо.

Синтезируемый компрессор должен сочетать достоинства кодеров на основе преобразований и субполосных кодеров. Для обеспечения возможности кодирования не только широкополосных речевых сигналов, но и передачи характеристик окружающей обстановки (среды в которой находится диктор) и аудиосигналов необходимо использовать многополосное возбуждение в составе CELP-кодера.

С целью минимизации количества кодируемых критических частотных полос необходимо осуществить их объединение в субполосы, обрабатываемые и кодируемые компрессором (см. рис. 11.1).

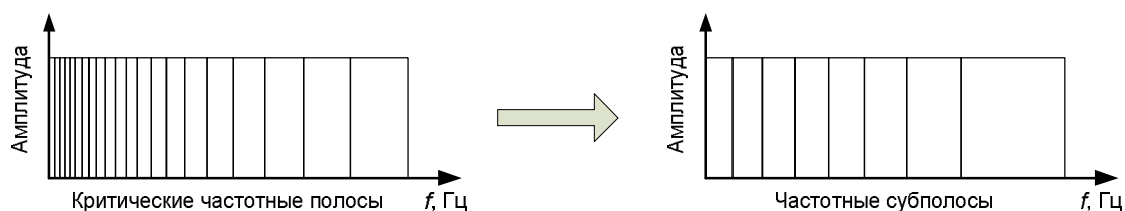


Рис. 11.1 – Объединение критических частотных полос в субполосы

При этом количество субполосных кодовых книг должно точно соответствовать количеству выбранных частотных субполос и определяться не только частотной избирательностью слуховой системы человека, но и формантной структурой речи. Кроме того, данные субполосы должны быть равнозначны в перцептуальном и статистическом смысле. Этого можно достичь, если применить для обучения кодовых книг субполосные векторы, полученные путем полосовой фильтрации широкополосного сигнала при помощи банка фильтров. Данный подход позволит смоделировать частотно-временной план, получаемый банком фильтров, т.е. сохранить образ банка фильтров в субполосных кодовых книгах. Компромисс между частотным и временным разрешением можно получить при помощи многоуровневых кодовых книг путем регулирования степени их детализации, которая будет отвечать за временное разрешение. Таким образом, задача получения адаптивного банка фильтров с перфективной реконструкцией преобразуется в задачу синтеза структуры мультиполосной многоуровневой кодовой книги форм сигналов возбуждения (см. рис. 11.2).

Для обучения книги можно использовать цифровой банк фильтров любой сложности, так как он применяется один раз на стадии проектировании банка кодовых книг (в так называемом «off-line» режиме).

При таком подходе полосно-отфильтрованные векторы будут близко ортогональными, и отпадет необходимость в банке фильтров с перфективной реконструкцией. Кроме того, если использовать квантование сигнала в порядке перцептуальной важности субполос, то возможна минимизация ошибки квантования предыдущей субполосы при квантовании следующей. При

этом каждый уровень квантования будет обеспечивать постепенную максимизацию перцептуально взвешенной меры пологости спектра и устранение эффекта «просачивания» энергии в соседние субполосы.

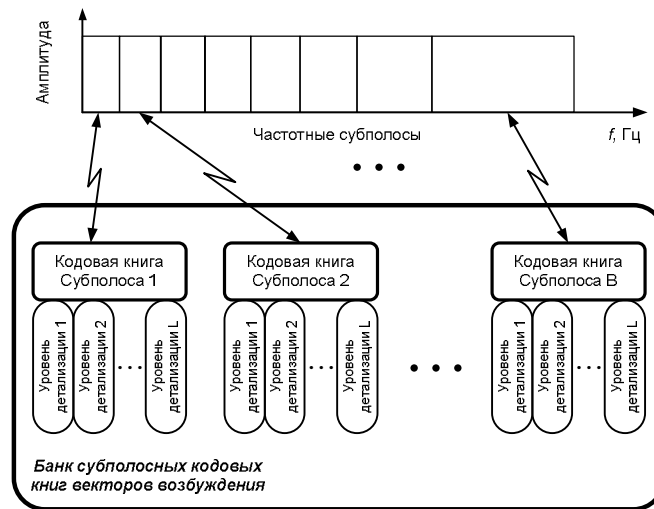


Рис. 11.2 – Отображение субполосной декомпозиции исходного сигнала на структуру банка кодовых книг субполосных сигналов возбуждения

Таким образом, задача может быть формализована в следующем виде. Пусть имеется множество обрабатываемых (кодируемых) субполос B с мощностью равной $C(B)$, кодер имеет $C(B)$ -полосную кодовую книгу с длиной субполосных векторов, равной L отсчетов. Длина обрабатываемого фрейма сигнала равна N отсчетов, длина векторов книги соответствует длине субфреймов анализа долговременного фильтра-анализатора (LTP). Количество субфреймов равно λ_c . Номер субфрейма λ . Индексы коэффициентов усиления субполосных кодовых векторов b -й субполосы для всех субфреймов определим как вектор $g_b = \{g_b^{(1)}, g_b^{(2)}, \dots, g_b^{(\lambda_c)}\}$, а индексы самих векторов возбуждения как $i_b = \{i_b^{(1)}, i_b^{(2)}, \dots, i_b^{(\lambda_c)}\}$. Обозначим набор индексов оптимальных коэффициентов усиления и субполосных векторов возбуждения как $g_{b,opt} = \{g_{b,opt}^{(1)}, g_{b,opt}^{(2)}, \dots, g_{b,opt}^{(\lambda_c)}\}$ и $i_{b,opt} = \{i_{b,opt}^{(1)}, i_{b,opt}^{(2)}, \dots, i_{b,opt}^{(\lambda_c)}\}$, соответственно. Тогда синтезированный субполосный сигнал для субфрейма λ и его коэффициент усиления определяются как функции от соответствующих индексов: $\hat{s}_{wb}(i_b^{(\lambda)})$ и $G_b(g_b^{(\lambda)})$.

Если обозначить фрейм входного сигнала длиной N отсчетов как $s(n)$, а субфрейм λ длиной L отсчетов – $s^{(\lambda)}(l)$, то $s(n) = \{s^{(1)}(l), s^{(2)}(l), \dots, s^{(\lambda_c)}(l)\}$. Фрейм взвешенного оригинального сигнала можно представить как набор субфреймов $\tilde{s}_w(n) = \{\tilde{s}_w^{(1)}(l), \tilde{s}_w^{(2)}(l), \dots, \tilde{s}_w^{(\lambda_c)}(l)\}$, операцию взвешивания через оператор $LP[\cdot]$, вектор ошибки квантования предыдущей субполосы как $e_{w(b-1)}(n) = \{e_{w(b-1)}^{(1)}(l), e_{w(b-1)}^{(2)}(l), \dots, e_{w(b-1)}^{(\lambda_c)}(l)\}$, остаточный сигнал после кодирования текущей субполосы представить $e_{wb}(n) = \{e_{wb}^{(1)}(l), e_{wb}^{(2)}(l), \dots, e_{wb}^{(\lambda_c)}(l)\}$, то процесс определения оптимального субполосного сигнала возбуждения может быть записан в следующем виде:

$$\tilde{s}_w(n) = LP[s(n)], n = 1, 2, \dots, N, 1 \leq \lambda \leq \lambda_c,$$

$$e_{w_0}^{(\lambda)} = \tilde{s}_w(k), l = 1, 2, \dots, L, k = (\lambda - 1) \cdot L + 1, (\lambda - 1) \cdot L + 2, \dots, \lambda \cdot L,$$

$$\{g_{b,opt}^{(\lambda)}, i_{b,opt}^{(\lambda)}\} = \arg \min_{g_b^{(\lambda)}, i_b^{(\lambda)}} e_{w_b}^{(\lambda)}(l) = e_{w_{(b-1)}}^{(\lambda)}(l) - G_b(g_b^{(\lambda)}) \cdot \hat{s}_{w_b}(i_b^{(\lambda)}), 1 \leq b \leq C(B), \quad (11.1)$$

$$e_w^{(\lambda)}(l) = e_{w_0}^{(\lambda)}(l) - \sum_{b=1}^{C(B)} G_b(g_{b,opt}^{(\lambda)}) \cdot \hat{s}_{w_b}(i_{b,opt}^{(\lambda)}),$$

$$e_w(n) = \{e_w^{(1)}(l), e_w^{(2)}(l), \dots, e_w^{(\lambda_c)}(l)\},$$

где $e_w(n)$ – остаточный сигнал (ошибка) после кодирования текущего фрейма взвешенного оригинального сигнала $\tilde{s}_w(n)$.

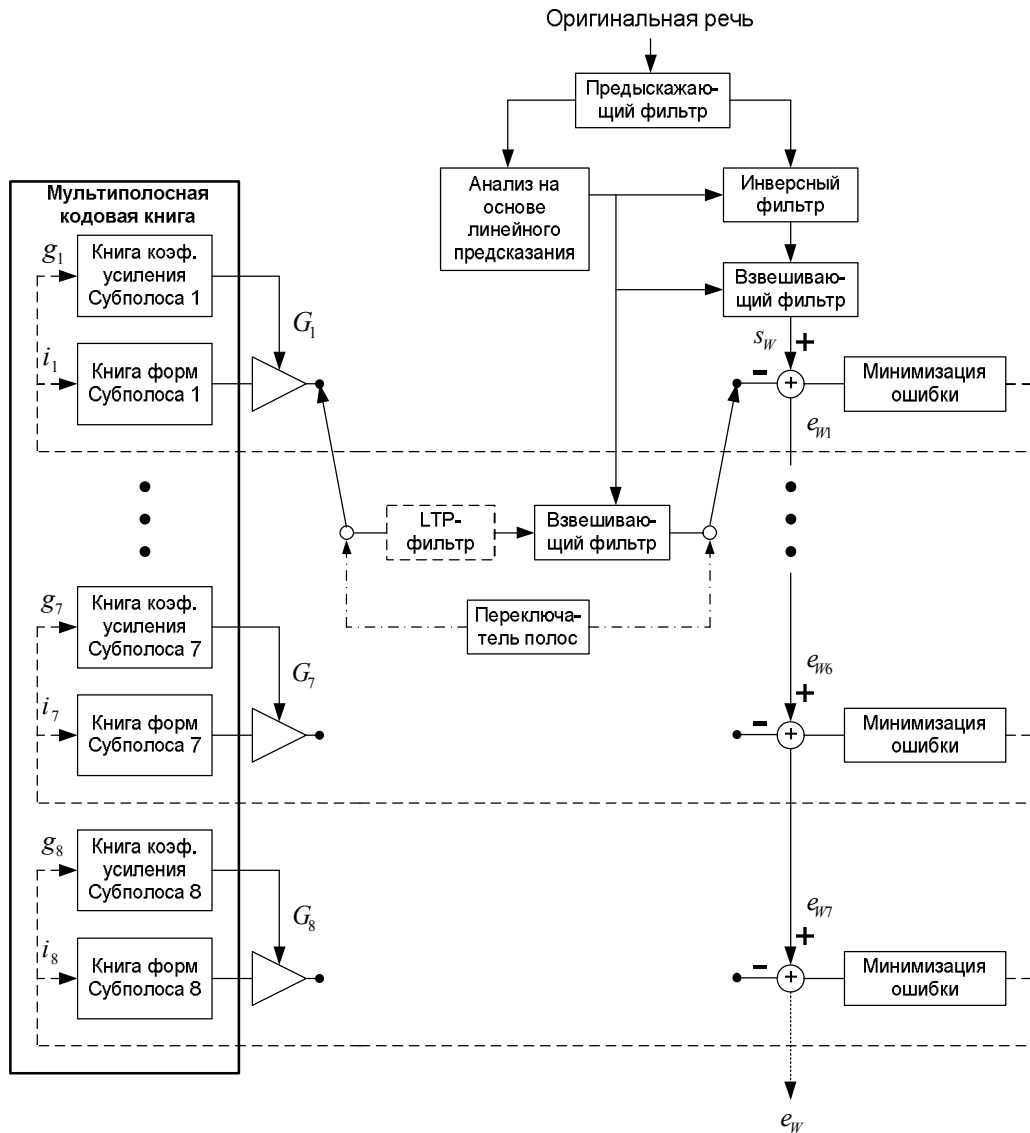


Рис. 11.3 – Структура перцептуального широкополосного кодера с мультиполосным возбуждением

Таким образом, обобщенная концепция кодера может быть представлена в виде структуры, проиллюстрированной на рис. 11.3.

Для того чтобы обеспечить перцептуальное управление структурой кодера и банком кодовых книг необходимо наличие соответствующей психоакустической модели, отражающей поведение слуховой системы человека и позволяющей статистически равнозначно осуществлять анализ и расчет порогов маскирования в критических частотных полосах. При этом встает вопрос о согласовании психоакустической модели с принятой схемой декомпозиции частотного диапазона на субполосы и, следовательно, в выборе соответствующего частотно-временного преобразования как базиса для расчета субполосных порогов маскирования.

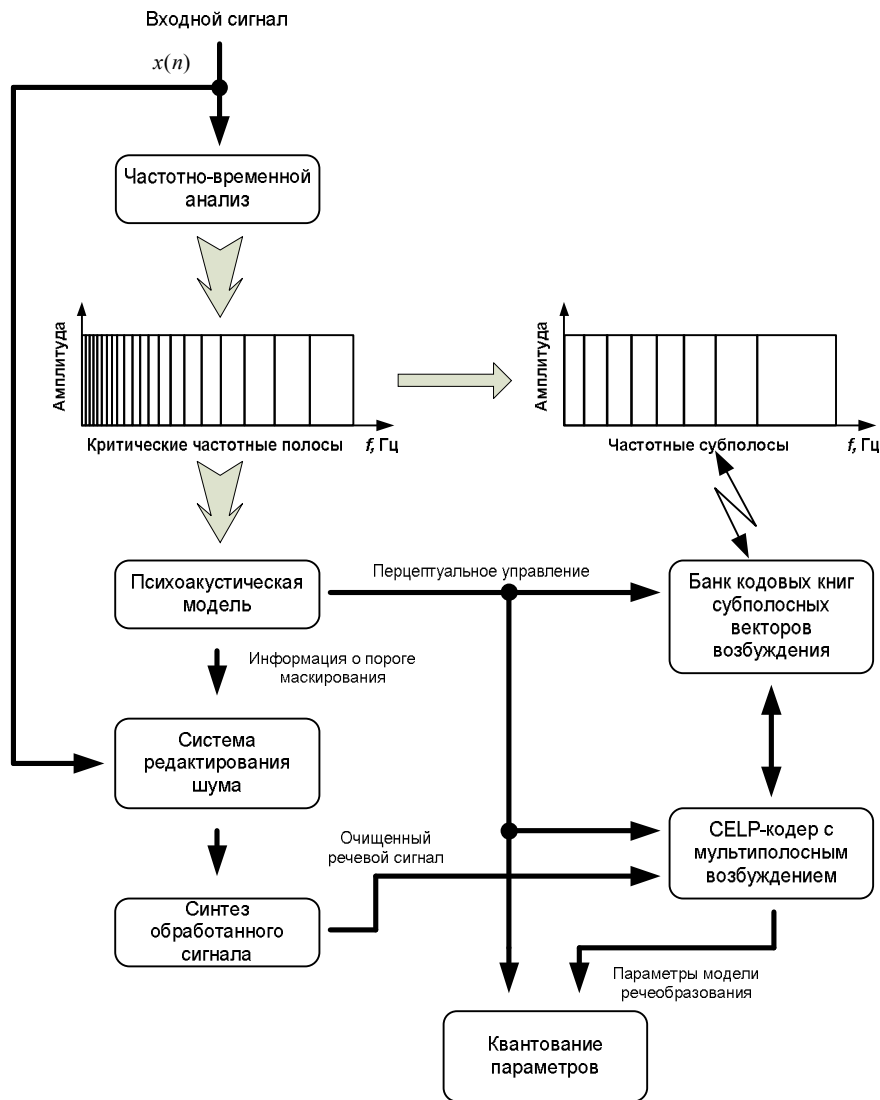


Рис. 11.4 – Модель комбинированной системы кодирования широкополосных речевых сигналов

Данная модель должна так же применяться на протяжении всей стадии кодирования и квантования коэффициентов модели речеобразования и параметров кодера. Так как системы мультимедиа применяются практически повсеместно. Большинство приложений связанных с кодированием и передачей речевых сообщений – это мобильные телефоны, информационные киоски, различные справочные службы и т.п., которые в большинстве случаев находятся в об-

ществленных местах в достаточно зашумленной обстановке и при этом должны обеспечивать высокое качество, разборчивость и узнаваемость речи. Для повышения работоспособности подобных систем в условиях окружающих шумов необходим препроцессор повышения качества речи, который должен гармонично сочетаться с архитектурой компрессора речевых сигналов.

Существует два возможных варианта объединения препроцессора редактирования шума и кодера речевых сигналов. Первый подразумевает простое каскадное соединение двух систем. Второй нацелен на более глубокую интеграцию и совместное использование некоторых модулей, как системой редактирования шума, так и кодером. При таком подходе результирующая система может быть более эффективно оптимизирована и обеспечит наилучшее качество.

Перцептуальный широкополосный компрессор и система редактирования шума должны работать согласованно, а, следовательно, все оценки и расчеты должны осуществляться в единой перцептуальной области на базе общей психоакустической модели и частотно-временного преобразования без дополнительных преобразований шкал. Схема комбинированной системы представлена на рис. 11.4.

11.2 Психоакустическая модель для оценки перцептуальной значимости субполос

11.2.1 Формирование субполосных групп на основе анализа их перцептуальной значимости

Базовый принцип субполосной декомпозиции заключается в выборе компромисса между следующими критериями: количеством субполос, их шириной (в каждую субполосу должно попадать приблизительно равное количество энергии), максимумом выигрыша при субполосном кодировании [1], простотой реализации банка фильтров. Кроме этого необходимо разбиение на неравные субполосы с целью отражения свойств слуховой системы человека [2,3] и формантной структуры речи [4]. Все современные кодеры на основе преобразования кодируют сигнал в частотной области в перцептуальной шкале. Для использования высокой степени сжатия линейного предсказания и перцептуальных возможностей кодеров на основе преобразования в CELP-кодере с мультиполосным возбуждением необходимо использовать многополосные кодовые книги, отражающие критические частотные полосы и обеспечивающие возможность подстройки структуры к перцептуальным свойствам кодируемого сигнала.

Проанализируем речевой сигнал и его структуру в частотной области. Для этого сравним два фрагмента: сигнал с мужскими и женскими голосами, и музыкой. Данные представлены на рис. 11.5.

Сравнение аудиосигналов, содержащих речевые сообщения с мужскими и женскими голосами показывает, что для мужских голосов характерен резкий спад амплитуды спектра мощности на частотах более 4000 Гц, в то время как для женских голосов наблюдается некоторая пологость спектра мощности в данной области и плавный спад к частоте 5000 Гц. Таким образом, для данных сигналов основная энергия сосредоточена в частотном диапазоне от 0 до 5000 Гц, а оставшийся диапазон частот содержит шумовую составляющую. Типичный спектр

музыкального фрагмента, показывает, что спектральные пики с наибольшей энергией появляются вплоть до частоты 6000 Гц, после чего наблюдается спад энергии. Как упоминалось выше, при разбиении на полосы необходимо учитывать не только специфику человеческого восприятия, но и формантную структуру человеческой речи. Остановимся на речевом сигнале и определим его параметры.

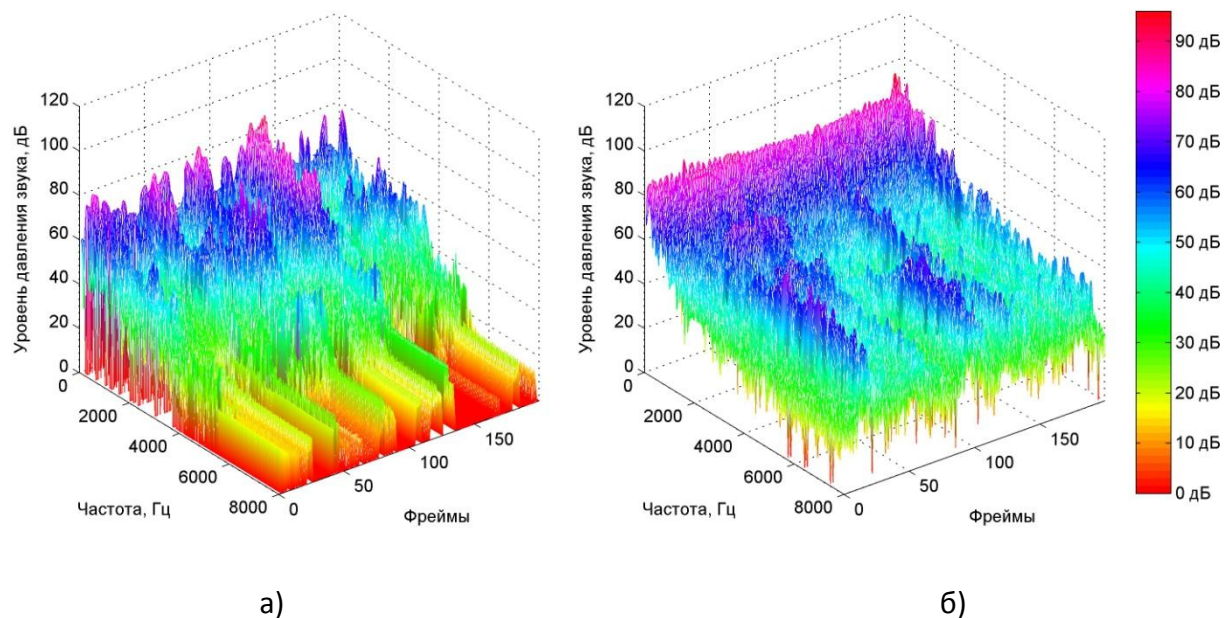


Рис. 11.5 – Анализ речевого сигнала с голосами дикторов-мужчин и дикторов-женщин (а) и музыкального фрагмента аудиосигнала (б) в частотной области

Табл. 11.1 – Параметры разбиения частотной шкалы на субполосы

№ субполосы	Диапазон, Гц	Кол-во гармоник	Значения гармоник, Гц	Кол-во барков
1	100-510	1	500	4
2	510-1080	1	1000	4
3	1080-1720	1	1500	3
4	1720-2320	1	2000	2
5	2320-3150	2	2500, 3000	2
6	3150-4100	2	3500, 4000	1.5
7	4100-5300	2	4500, 5000	1.5
8	5300-7700	5	5500, ..., 7500	2

Широко известный факт, что речевой сигнал имеет частоту основного тона, которая, в зависимости от речевого аппарата человека, может колебаться и находится в диапазоне от 100 до 500 Гц [4]. При этом более низкие значения характерны для мужских голосов, более высокие – для женских (вплоть до 400 Гц), наивысшее значение частоты основного тона характерно для детей. Учитывая базовые принципы CELP-модели (в части метода оценки частоты основного тона), можем заключить, что для корректной оценки последней в первую полосу должны попадать частоты, соответствующие диапазону варьирования основного тона. Так как речь человека имеет ярко выраженную формантную структуру, то все последующие полосы необхо-

димом группировать таким образом, чтобы в них попадали все возможные высшие гармоники частоты основного тона. Если принять, что наивысшая частота основного тона ребенка достигает $f_0 = 500$ Гц, и ее высшие гармоники будут определяться рядом $2f_0, 3f_0, \dots, nf_0$, то, учитывая формантную структуру речи и критерии человеческого восприятия, можно сгруппировать 21 критическую частотную полосу в субдиапазоны (субполосы). При этом необходимо учесть, что основная информация (энергия) сосредоточена в диапазоне частот до 5000 Гц, следовательно, необходимо чтобы в максимально возможное количество субдиапазонов попадало по одной гармонике частоты основного тона. На основании всего вышесказанного сформируем субдиапазоны (субполосы), параметры которых представлены в табл. 11.1.

Проанализируем субполосные группы на примере СПМ типового фрейма речевого сигнала (см. рис. 11.6).

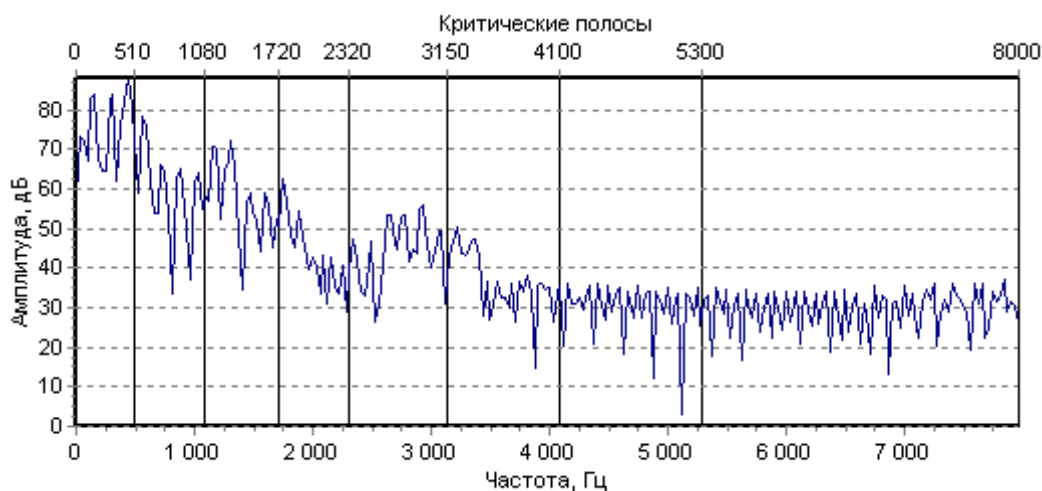


Рис. 11.6 – Спектральная плотность мощности фрейма речевого сигнала

Анализ рис. 11.6 показывает, что принятое разбиение на субполосы, учитывающее критерий человеческого восприятия и формантную структуру речи, позволяет довольно качественно осуществлять анализ (форманты речевого фрейма, точно попадают в принятые субполосы). Распределение энергии в выбранных субполосах для широкополосного речевого и музыкального сигнала представлены на рис. 11.7.

Как видно из рис. 11.7, для всех типов аудиоматериала в первых трех субполосах сосредоточена большая часть энергии сигнала. Данные субполосы являются наиболее важными (информативными) с перцептуальной точки зрения, в то время как все остальные субполосы имеют приблизительно равное распределение энергии.

В качестве частотно-временного преобразования в блоке психоакустического анализа в кодере (см. рис. 11.4) может использоваться: банк цифровых фильтров (БЦФ) (тот же, что и при обучении кодовых книг), ДПФ или WDFT [5-7].

Применение БЦФ с неравными полосами обусловлено сложностью получения высококачественного банка с хорошим разделением каналов (с отсутствием эффекта «просачивания» энергии между соседними полосами). При этом «просачивание» будет вносить существенную ошибку в оценку субполосных порогов маскирования. Кроме того, реализация имеет достаточно высокую вычислительную сложность.

ДПФ позволяет осуществлять анализ с равномерным разрешением при этом энергия низкочастотных полос, в которых сосредоточена основная информация, определяется по 1–3 коэффициентам, в то время как на менее информативные высокочастотные полосы приходится на порядок больше коэффициентов. Преимуществом данного подхода является простота реализации и небольшая вычислительная сложность (доступны быстрые алгоритмы). Недостатком является весьма грубая оценка порогов маскирования.

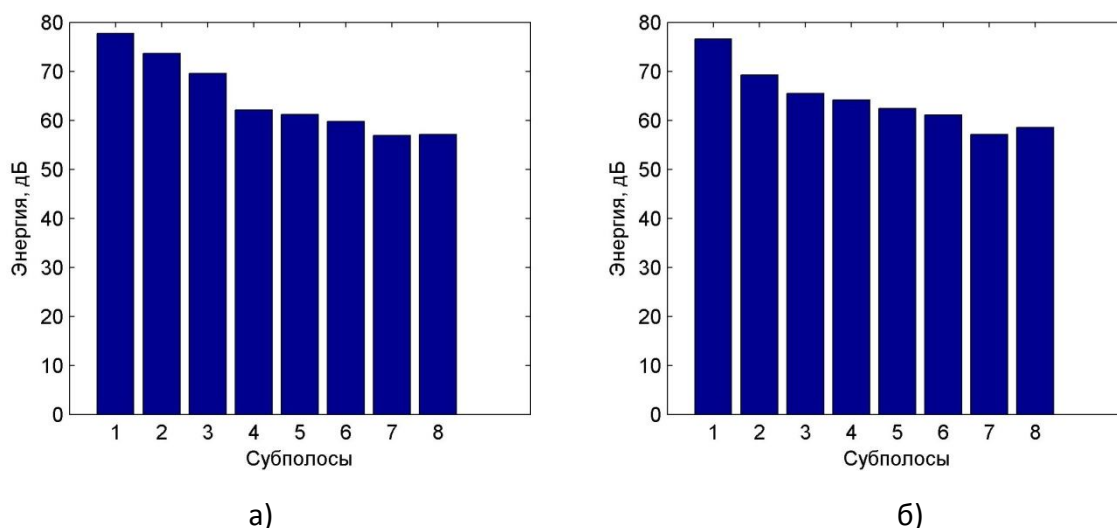


Рис. 11.7 – Распределение энергии в выбранных субполосах: а) речевая база данных, б) музыкальная база данных

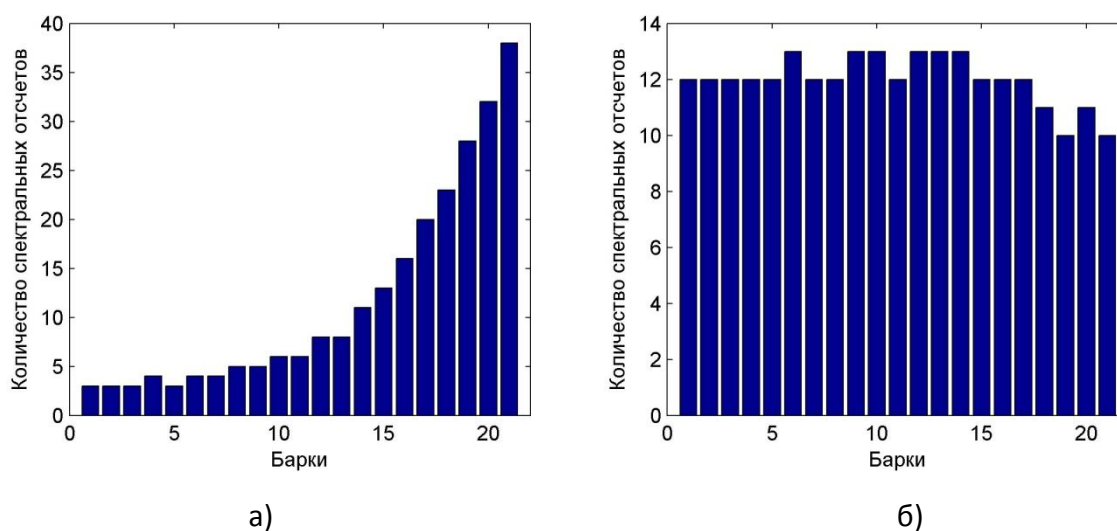


Рис. 11.8 – Распределение коэффициентов преобразования по критическим частотным полосам: а) ДПФ, б) WDFT

Применение же WDFT позволяет деформировать ось частот в соответствии с принятой психоакустической шкалой, в результате анализ во всех критических частотных полосах осуществляется статистически равноценно (каждая полоса характеризуется равным количеством коэффициентов преобразования). Данный подход достаточно прост в реализации, однако не имеет быстрых алгоритмов для вычисления, сравнимых с БПФ. Несмотря на это он сочетает в себе достоинства предыдущих двух методов.

Таким образом, в качестве частотно-временного преобразования для психоакустической модели необходимо использовать WDFT, которое обеспечит статистически равнозначный анализ энергии сигнала в критических частотных полосах (Барках) (см. рис. 11.8).

11.2.2 Оценка порогов маскирования в WDFT-области

В известной работе Джонсона [8] психоакустическая модель основана на ДПФ: расчет ДПФ взвешенного сегмента сигнала, группировка коэффициентов преобразования в группы, соответствующие критическим частотным полосам восприятия, и расчет энергии в критических частотных полосах. Достижение приемлемого спектрального разрешения в наименьших критических частотных полосах требует использования ДПФ с достаточно длинным временным окном. Поэтому концептуальная простота и эффективность нивелируются недостаточным временным разрешением, неприемлемым для анализа более тонкого феномена, такого как маскирование назад («pre-masking») [9]. Второй класс, использующий банки фильтров для декомпозиции сигнала на неравномерные полосы, исключает этот недостаток, однако имеет достаточно высокую вычислительную сложность, особенно если необходима хорошая аппроксимация критических полос. Ни одно из решений не превосходит другое, оба из подходов находят свое применение.

Исследования, проведенные в работах [6,10,11] показали, что малоразмерное WDFT, может успешно заменить ДПФ с большой длиной выборки. Это оказалось возможным благодаря тому, что WDFT позволяет разместить частотные компоненты в соответствии с распределением критических частотных полос, поэтому в психоакустической модели на базе WDFT могут быть уравновешены как хорошее частотное, так и временное разрешение.

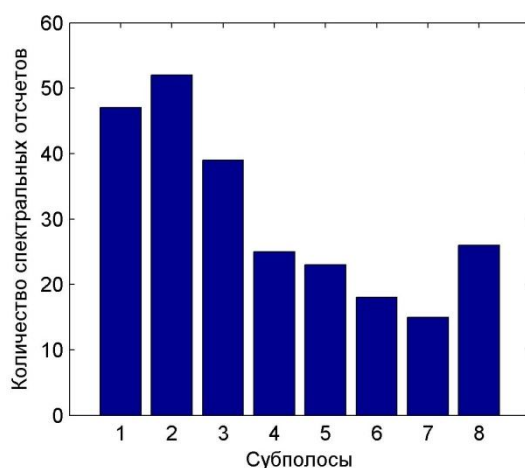


Рис. 11.9 – Распределение сгруппированных коэффициентов преобразования

Для использования свойств WDFT процедура оценки порога маскирования на базе общей психоакустической модели Джонсона была модифицирована и подстроена под широкополосный речевой сигнал с полосой 8 кГц. Основные шаги оценки порога маскирования следующие: расчет энергии критических частотных полос по WDFT-спектру мощности; свертка с функцией распространения; вычитание относительного смещения порога; нормализация и сравнение с абсолютным порогом слышимости.

Энергия рассчитывается в критических частотных полосах путем суммирования компонент спектра мощности в каждой критической частотной полосе. Коэффициенты преобразования группируются в субполосы в соответствии с табл. 11.1, в каждой из которых осуществляется оценка порогов маскирования. Распределение сгруппированных коэффициентов преобразования представлено на рис. 11.9.

Дискретное представление энергии критической частотной полосы может быть выражено как взвешенная сумма компонент спектра мощности:

$$E_b = \frac{1}{2\pi} \sum_{k \in K_b} \Delta\omega_k |\hat{S}(\omega_k)|^2, \quad (11.2)$$

где b – номер критической частотной полосы; K_b – количество коэффициентов преобразования, попадающих в b -ю критическую частотную полосу; k – индекс коэффициента преобразования (спектрального отсчета); $\hat{S}(\omega_k)$ – оценка спектра оригинального речевого сигнала.

На следующем шаге вычисляется свертка энергий критических частотных полос E_b с функцией распространения по базилярной мембране SF_b для учета распространения маскирования:

$$C_b = E_b * SF_b. \quad (11.3)$$

Относительное смещение порога в каждой критической частотной полосе рассчитывается с использованием меры тональности. Для определения типа маскера («шумоподобный» или «тональный») используется мера пологости спектра (от англ. – Spectral Flatness Measure – SFM), которая определяется как отношение среднего геометрического μ_{gb} к среднему арифметическому μ_{ab} спектральных компонент внутри критической полосы b :

$$SFM_b = 10 \log_{10} \left(\frac{\mu_{gb}}{\mu_{ab}} \right), \text{ [дБ]}. \quad (11.4)$$

Данная оценка может быть затруднена в случае обычных моделей с малым разрешением, так как количества ДПФ-коэффициентов в низкочастотных полосах обычно не достаточно для эффективного расчета SFM . В подобных случаях SFM рассчитывается для всего спектра или predeterminedных тональностей для каждой критической частотной полосы. В случае $WDFT$, коэффициенты преобразования группируются равномерно в критических частотных полосах, следовательно, SFM может быть определена для каждой полосы отдельно.

Смещение порога маскирования O_b оценивается по следующему выражению:

$$O_b = ton_b \cdot (14.5 + b) + (1 - ton_b) \cdot 5.5, \text{ [дБ]}, \quad (11.5)$$

где ton_b – индекс тональности, определяемый как

$$ton_b = \min \left(\frac{SFM_b}{-60 \text{ дБ}}, 1 \right). \quad (11.6)$$

Для получения энергии порога маскирования относительное смещение (11.5) вычитается из свернутого спектра критической полосы (11.3):

$$P_{TT,b} = 10^{(\log_{10} C_b - O_b / 10)}, \quad (11.7)$$

а затем осуществляется нормализация и сравнение с абсолютным порогом слышимости.

Пример оценки порогов маскирования в критических частотных полосах и субполосах представлен на рис. 11.10.

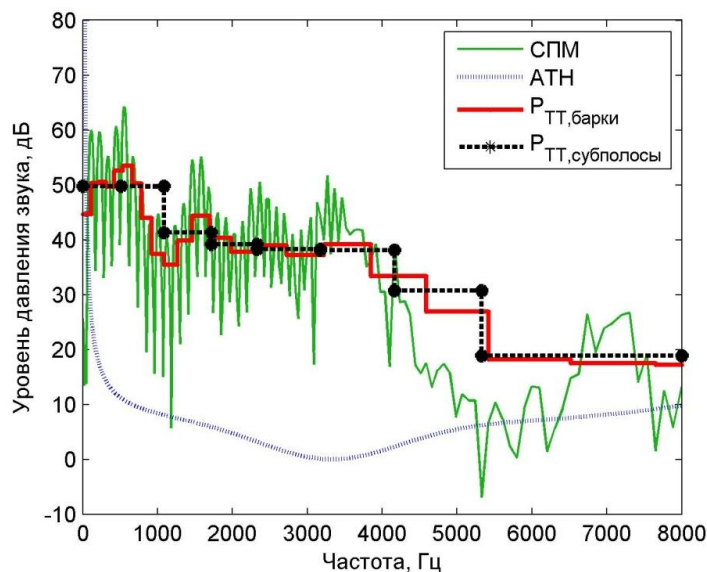


Рис. 11.10 – Пример оценки субполосных порогов маскирования

Анализ рис. 11.10 показывает, что оценки порогов маскирования в критических частотных полосах и в выбранных субполосах имеют приблизительно одинаковое значение. Таким образом, психоакустическая модель, построенная на базе WDFT, обеспечивает статистически равнозначный анализ энергии сигнала в субполосах. Кроме того, данная модель может быть использована для более гибкого управления процессом кодирования сигнала. Применение динамической оценки энергии субполосных порогов маскирования для каждого фрейма кодируемого аудиосигнала позволит реконфигурировать банк кодовых книг.

11.3 Математическая модель перцептуального CELP-кодера с реконфигурируемым банком кодовых книг

Для реконфигурации банка кодовых книг необходимо определение перцептуальной важности всех субполос или субполосной перцептуальной энтропии, которое осуществляется по оценке энергии субполосного порога маскирования [12]:

$$SPE_b = \log_2 \left(\frac{E_b}{\sqrt{\frac{6 \cdot P_{TT,b}}{bandwidth_b}}} + 1 \right), \left[\frac{\text{бит}}{\text{полосу}} \right] \quad (11.8)$$

где $bandwidth_b$ – ширина субполосы b в спектральных отсчетах.

Данная оценка позволяет определить порядок обработки (кодирования) субполос по мере их значимости. Кроме того, субполосную перцептуальную энтропию можно использовать для организации эффективного управления степенью детализации в субполосе [12,13].

Формализуем алгоритм трансформации банка кодовых книг.

Пусть имеется банк кодовых книг с множеством уровней детализации каждой субполосной книги $CBD_b = \{CBD_b^1, CBD_b^2, \dots, CBD_b^{l_b}\}$, где l_b – максимально возможная глубина поиска в субполосной кодовой книге (количество уровней детализации). Для того чтобы преобразовать перцептуальную энтропию SPE_b каждой субполосы в глубину поиска d_b (номер уровня детализации) в соответствующей субполосной кодовой книге, вычисляется вклад каждого уровня книги DPE_b в суммарное значение субполосной перцептуальной энтропии [12]. Кроме того, в процессе кодирования участвуют только субполосы с перцептуальной энтропией больше нуля, то есть формируется подмножество кодируемых субполос $Bands \in \mathbb{B}_{SPE_b > 0}$. С учетом введенных обозначений алгоритм преобразования SPE_b в d_b может быть представлен в виде выражения (11.9):

$$SPE_{max} = \max(SPE),$$

$$DPE_b = \begin{cases} \frac{SPE_{max}}{l_b}, & \text{если } l_b \geq SPE_{max} \\ \text{иначе } \frac{l_b}{SPE_{max}} \end{cases}, b \in Bands, \quad (11.9)$$

$$d_b = \min(l_b, \text{nint}(SPE_b \cdot DPE_b)),$$

$$\text{depth}(i) = \begin{cases} d_i, & \text{если } i \in Bands \\ \text{иначе } 0 \end{cases}, i = 1, 2, \dots, C(B).$$

В результате своей работы алгоритм преобразования SPE возвращает вектор depth с номерами требуемых уровней детализации для каждой субполосной кодовой книги. На основании уточненной модели кодирования субполос и алгоритма динамической трансформации банка кодовых книг (11.9), упрощенную схему реконfigurирования структуры мультиполосной многоуровневой кодовой книги можно представить в виде, проиллюстрированном на рис. 11.11.

Принцип изменения структуры книги заключается в следующем. Оригинальный речевой сигнал (очередной кодируемый фрейм) поступает на вход блока WDFТ-анализа, в котором осуществляется оценка субполосных порогов маскирования $P_{TT,b}$. Затем производится оценка перцептуальной энтропии каждой из субполос кодера SPE_b . Таким образом, на выходе получается своего рода стратегия выделения бит или вектор SPE , содержащий значения перцептуальной важности кодируемых субполос для текущего фрейма сигнала. Данная стратегия, согласно алгоритму (11.9), преобразуется в вектор depth , содержащий глубину поиска (номер уровня детализации векторов возбуждения) в субполосных кодовых книгах. При этом если $SPE_b = 0$, то данная субполоса удаляется из рассмотрения, т.е. не кодируется индекс вектора возбуждения и его коэффициент усиления.

Процесс реконfigurации банка кодовых книг по этапам представлен на рис. 11.12. Обозначим прямое частотно-временное преобразование оператором D . Пусть алгоритм расчета порогов маскирования, включающий оценку субполосной перцептуальной энтропии, будет представлен оператором $MT[\cdot]$. Алгоритм (11.9) представим оператором $PE[\cdot]$, операцию сортировки субполос, участвующих в процессе кодирования, в порядке убывания SPE_b –

$descend[\cdot]$. Тогда, если применить взвешивание минимизируемой ошибки кодирования на субполосный порог маскирования $P_{TT,b}$, то уточненная математическая модель кодирования широкополосного речевого сигнала может быть выражена в виде, представленном выражениями (11.10), при этом синтезированный субполосный сигнал теперь будет являться функцией от индекса вектора возбуждения на заданном уровне детализации $\hat{s}_{w_b}(i_b^{(\lambda)}, depth(b))$.

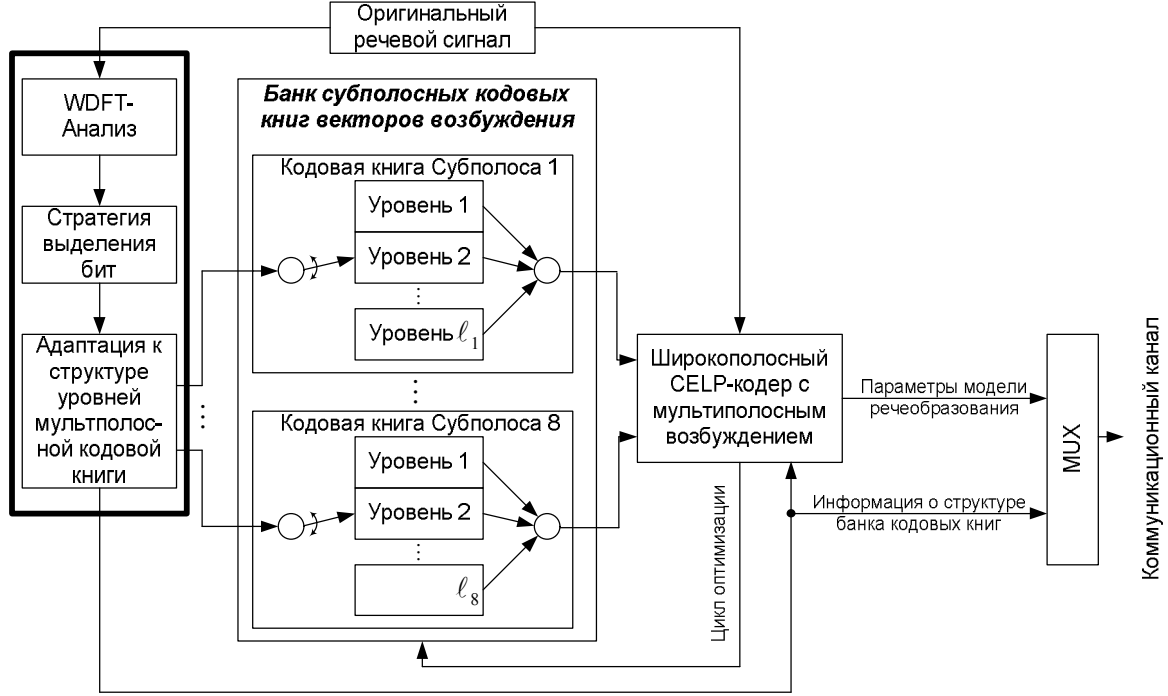


Рис. 11.11 – Схема реконфигурации структуры кодовой книги широкополосного кодера на базе WDFT-психоакустической модели

$$\{B, SPE, P_{TT}\} = \{D, MT[\cdot], PE[\cdot]\},$$

$$Bands \in descend \left[\begin{matrix} B \\ SPE > 0 \end{matrix} \right],$$

$$\tilde{s}_w(n) = LP[s(n)], n = 1, 2, \dots, N, 1 \leq \lambda \leq \lambda_c,$$

$$e_{w_0}^{(\lambda)} = \tilde{s}_w(k), l = 1, 2, \dots, L, k = (\lambda - 1) \cdot L + 1, (\lambda - 1) \cdot L + 2, \dots, \lambda \cdot L,$$

$$b = \overline{1, C(Bands)},$$

$$\{g_{b,opt}^{(\lambda)}, i_{b,opt}^{(\lambda)}\} = \arg \min_{g_b^{(\lambda)}, i_b^{(\lambda)}} e_{w_b}^{(\lambda)}(l) = \frac{e_{w_{(b-1)}}^{(\lambda)}(l) - G_b(g_b^{(\lambda)}) \cdot \hat{s}_{w_b}(i_b^{(\lambda)}, depth(b))}{P_{TT,b}}, \quad (11.10)$$

$$e_w^{(\lambda)}(l) = e_{w_0}^{(\lambda)}(l) - \sum_{b=1}^{C(B)} G_b(g_{b,opt}^{(\lambda)}) \cdot \hat{s}_{w_b}(i_{b,opt}^{(\lambda)}),$$

$$e_w(n) = \{e_w^{(1)}(l), e_w^{(2)}(l), \dots, e_w^{(\lambda_c)}(l)\}.$$

Таким образом, схема перцептуального широкополосного кодера речевых сигналов с реконфигурируемой структурой банка кодовых книг векторов возбуждения преобразуется к виду, представленному на рис. 11.13 и рис. 11.14.

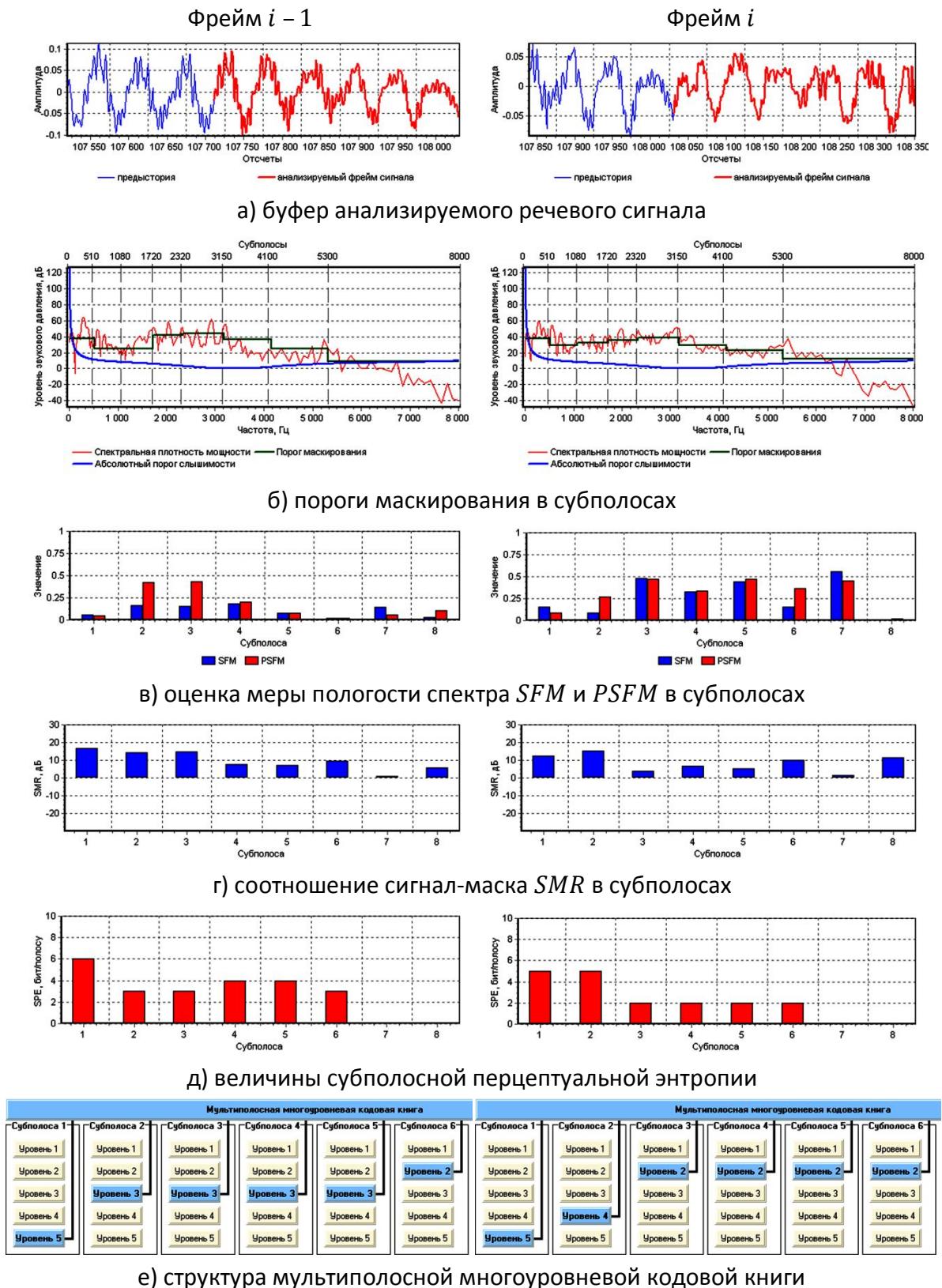


Рис. 11.12 – Пример работы психоакустической модели с результатами реконфигурации мультиполосной многоуровневой кодовой книги

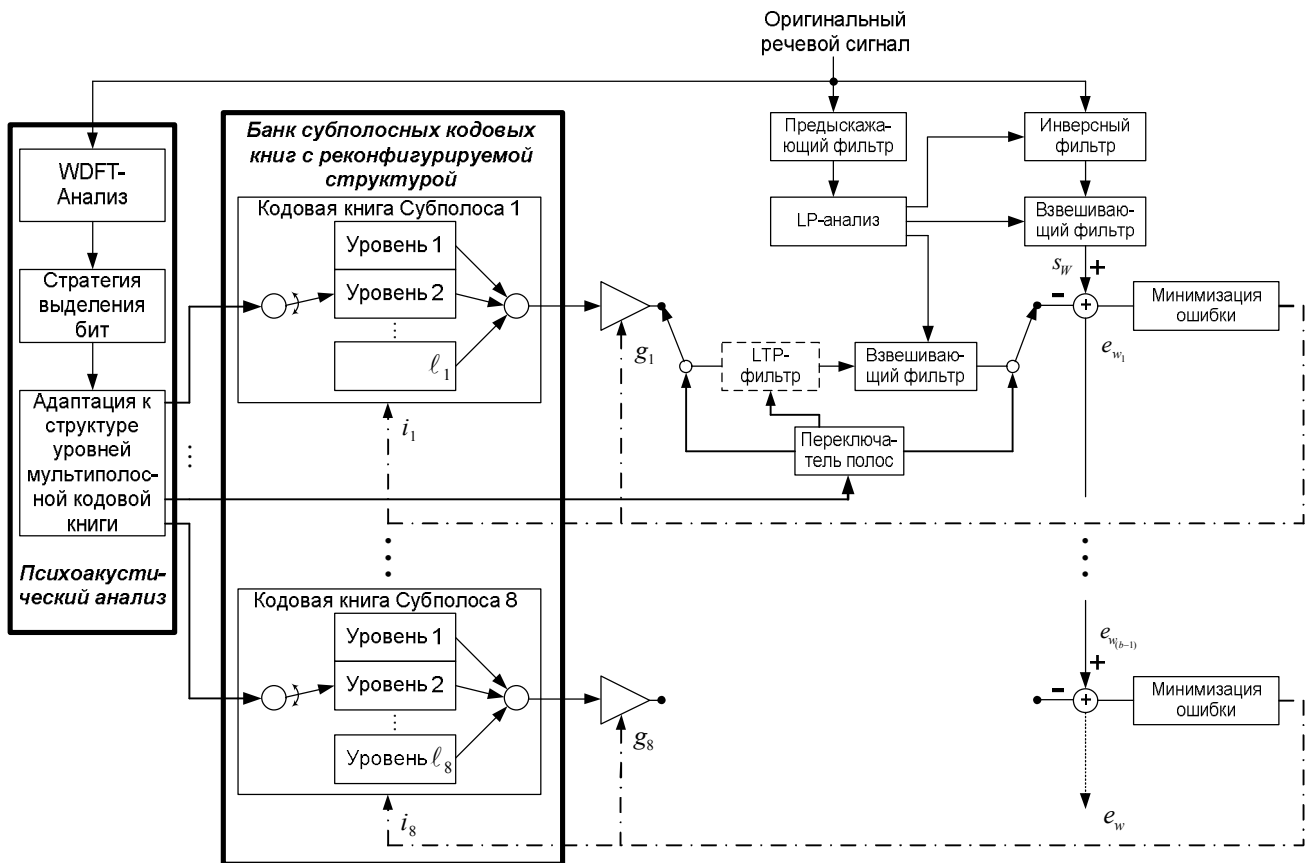


Рис. 11.13 – Схема кодера с реконфигурируемой структурой банка кодовых книг

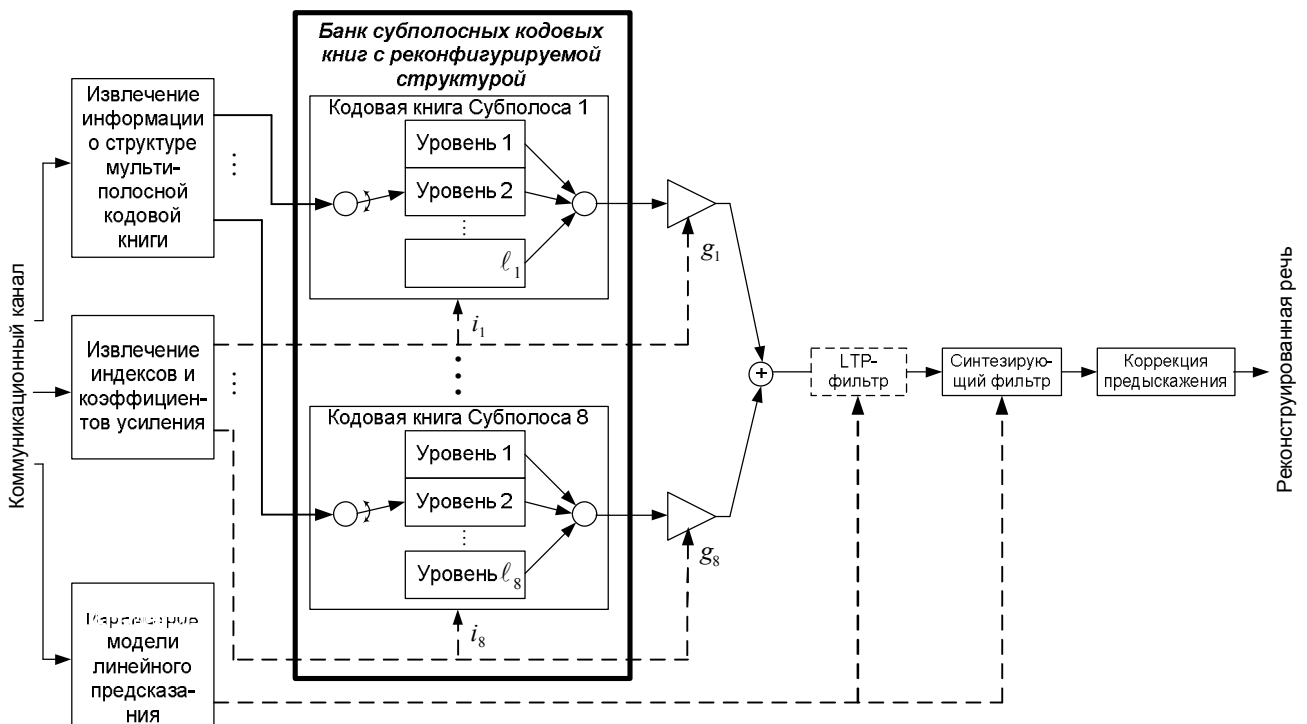


Рис. 11.14 – Схема декодера с реконфигурируемой структурой банка кодовых книг

11.3.1 Синтез субполосных кодовых книг с варьируемым уровнем детализации (количеством векторов)

В общем случае обучение кодовых книг на фрагментах музыкальных композиций с присутствующим вокальным сопровождением обеспечивает лучшее качество реконструированной речи не только по объективным оценкам, но и субъективным [14].

Для описываемого широкополосного кодера речевых сигналов банк фильтров нужен только для получения субполосных обучающих множеств с целью формирования банка кодовых книг векторов возбуждения. Для получения субполосных сигналов используется косинусно-модулированный банк фильтров с неравными субполосами [15], отражающими принятую схему декомпозиции (табл. 11.1). АЧХ данного банка цифровых фильтров представлена на рис. 11.15.

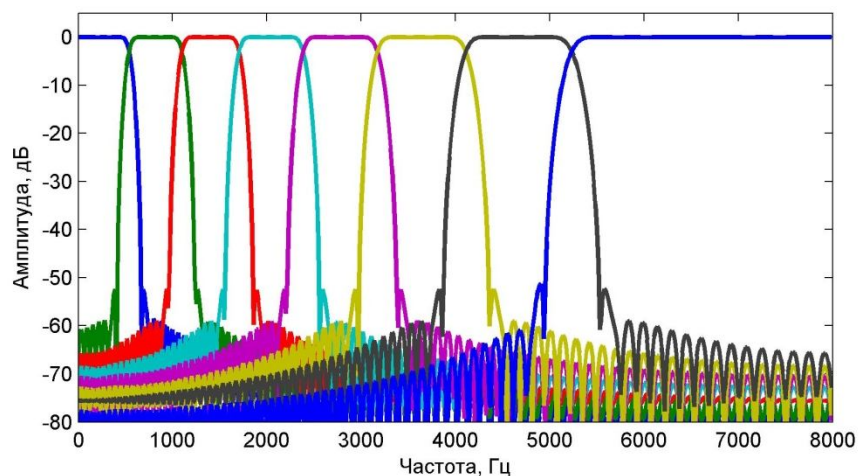


Рис. 11.15 – АЧХ банка фильтров, осуществляющего разбиение частотной оси на субполосы

Необходимо отметить, что отфильтрованные кодовые векторы имеют конечную длительность, равную длине субфрейма, принятого для осуществления LTP-анализа. Это обуславливает утечку («просачивание») энергии между смежными кодовыми книгами (субполосами). Однако «просачивание» спектральной энергии i -й субполосы может быть компенсировано при квантовании $i + 1$ -й субполосы, поэтому нет необходимости в банке фильтров с совершенной реконструкцией. Кроме того, сама модель линейного предсказания обладает некоторыми интегрирующими свойствами (синтезирующий фильтр окрашивает «шумоподобный» сигнал возбуждения), что позволяет скорректировать спектр ошибки при помощи огибающей спектра текущего фрейма кодируемого сигнала. Кроме того, так как кодовые векторы разных кодовых книг близко ортогональные, последовательный поиск в каждой субполосной кодовой книге обеспечивает почти тоже качество, что и оптимальный совместный поиск во всех субполосных книгах, но со значительным снижением сложности.

Для обучения кодовых книг векторов возбуждения использовался модифицированный алгоритм K -средних, не требующий задания порога (величины улучшения) общего искажения как признака завершения стадии обучения кодовой книги [16],

Увеличение предельной глубины субполосных кодовых книг свыше 256 векторов, не обеспечивает существенного прироста качества кодирования сигнала. В результате исследований было определено, что оптимальный порядок фильтра-предсказателя для частоты дискре-

тизации $f_s=16$ кГц составляет 16, длительность субвектора – 5 мс или 80 отсчетов, размер фрейма равен 320 отсчетов, а предельная глубина книги (количество уровней детализации) – 256 векторов.

Значительная вычислительная сложность поиска данных в кодовой книге обуславливает необходимость в банке субполосных кодовых книг, который бы обеспечивал не только высокое качество звучания реконструированного сигнала, но и эффективный поиск субполосных векторов возбуждения. Повышения скорости поиска оптимального вектора в кодовой книге с учетом заданного предельного уровня детализации можно добиться, если использовать структурированную кодовую книгу, в которой поиск оптимального вектора на предыдущем уровне является начальным условием (отправной точкой) поиска на следующем. Таким образом, пространство поиска постепенно уменьшается, и область оптимальных векторов локализуется (рис. 11.16).

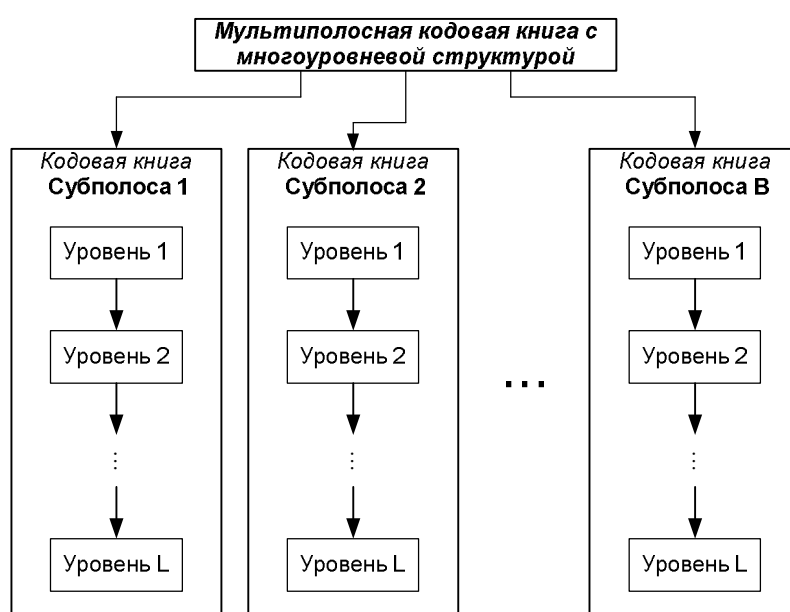


Рис. 11.16 – Зависимая структура многоуровневых кодовых книг

Такая структура банка кодовых книг обеспечит масштабируемость потока данных. При этом информация более высокого уровня детализации будет включать информацию обо всех более низких уровнях. Подобная кодовая книга должна иметь древовидную структуру с расщеплением на ветви каждого узла. Для формирования ее уровней (узлов дерева) необходимо задание соответствующих начальных условий и обучающих множеств. Процесс обучения такого банка субполосных кодовых книг можно представить следующим образом:

- 1) формирование исходного субполосного обучающего множества;
- 2) кластеризация всего обучающего множества на кластеры первого уровня детализации и формирование подмножеств первого уровня;
- 3) применение того же алгоритма обучения для каждого из подмножеств первого уровня для получения кодовых векторов и подмножеств второго уровня;
- 4) алгоритм повторяется для заданного количества уровней детализации.

Таким образом, вычислительная сложность поиска оптимального вектора возбуждения в древовидной кодовой книге будет существенно ниже, чем в книге с независимыми уровнями.

Кроме этого, качество такого банка сопоставимо с банком книг с независимой структурой уровней.

Выбранная структура банка кодовых книг векторов возбуждения позволяет пользователю в зависимости от требований, предъявляемых к качеству речи и пропускной способности коммуникационного канала, выбрать количество кодируемых субполос и ограничить глубину поиска в субполосных кодовых книгах. То есть существует возможность заранее задать предельную структуру банка. При этом частотно-временной план может быть как фиксированным (фиксированная структура банка кодовых книг), так и адаптивным, подстраиваемым под характеристики фрейма кодируемого сигнала (реконфигурируемая структура банка кодовых книг).

При квантовании структуры банка кодовых книг возможно использование трех подходов: применение кодов с минимальной избыточностью [17], задание конечного числа вариантов стратегий выделения бит на субполосы [18] и векторное квантование субполосной перцептуальной энтропии [14].

Анализ распределения субполосной энтропии для речевой базы и музыкальной базы показывает, что каждая субполоса имеет свой закон распределения. Таким образом, для кодирования перцептуальной энтропии при помощи кодов с минимальной избыточностью для каждой субполосы потребуется своя собственная кодовая книга. При этом количество возможных вариантов значений субполосной перцептуальной энтропии достаточно велико, следовательно, построение эффективного квантователя структуры банка кодовых книг на базе кодирования с минимальной избыточностью является неоптимальным.

Применение же подхода с фиксированным множеством вариантов распределения перцептуальной энтропии [18], с увеличением количества субполос имеет свои ограничения в плане кодирования варианта сочетания SPE в каждом конкретном случае. Например, при 8 субполосах количество комбинаций составит 40320.

Однако решение было найдено в работе [14], в которой предлагается квантовать структуру банка кодовых книг при помощи векторного квантователя по 10-битной кодовой книге. Данный подход позволил при минимальном количестве бит осуществить достаточно качественное квантование структуры банка кодовых книг с максимальным использованием свойств векторного квантования (взаимную корреляцию элементов вектора). Тем самым удалось построить книгу из наиболее вероятных структур банка для всего обучающего множества. Кроме того, такой подход обеспечивает качество кодирования сигнала сопоставимое с качеством, обеспечиваемым неквантованной реконфигурируемой структурой банка кодовых книг векторов возбуждения.

11.4 Препроцессор повышения качества речи

Большинство существующих систем подавления шума работают в частотной области, при этом используется хорошо известный подход спектрального вычитания [19,20]. Несмотря на то, что этот метод очень прост и легко реализуется, его слабым местом является остаточный шум, также известный как «музыкальный тон». Следовательно, возникает задача модификации взвешивающего правила таким образом, чтобы оставить «музыкальный тон» немного ниже по-

рога маскирования [21]. Основными элементами препроцессора повышения качества речи является блок частотно-временного преобразования (используется для анализа-синтеза) и психоакустическая модель. С их помощью осуществляется оценка энергии сигнала и порогов маскирования в частотной области, а также перцептуальное взвешивание для устранения эффекта «музыкального тона».

Так как в описываемой системе компрессии широкополосного речевого сигнала уже выбран тип частотно-временного преобразования и используется психоакустическая модель для управления реконfigurацией структуры банка кодовых книг, то с целью экономии вычислительного ресурса и достижения высокого качества комбинированной системы необходимо использовать второй подход (обе системы используют общие ресурсы: частотно-временное преобразование, психоакустическая модель и т.п.). Таким образом, общая структура комбинированной системы кодирования широкополосной речи может быть представлена в виде схемы, изображенной на рис. 11.17.

Применение WDFT в перцептуальной системе редактирования шума не только как базиса для определения модели маскирования, но и как инструмента спектрального анализа, позволит добиться наилучшего качества по сравнению с обычными системами повышения качества речи. Это объясняется тем, что весь процесс обработки осуществляется в единой перцептуальной области (нет необходимости в преобразованиях между различными частотными шкалами). Более того, обработка речевого сигнала, осуществляемая в критических частотных полосах, более точна в контексте психоакустического моделирования.

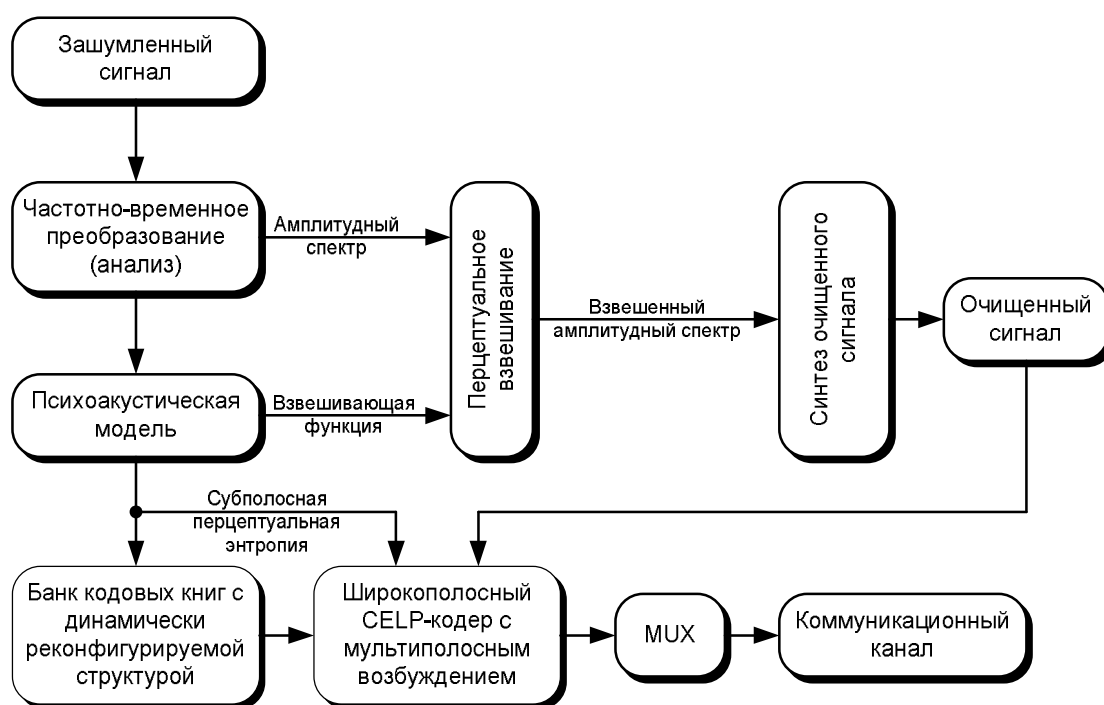


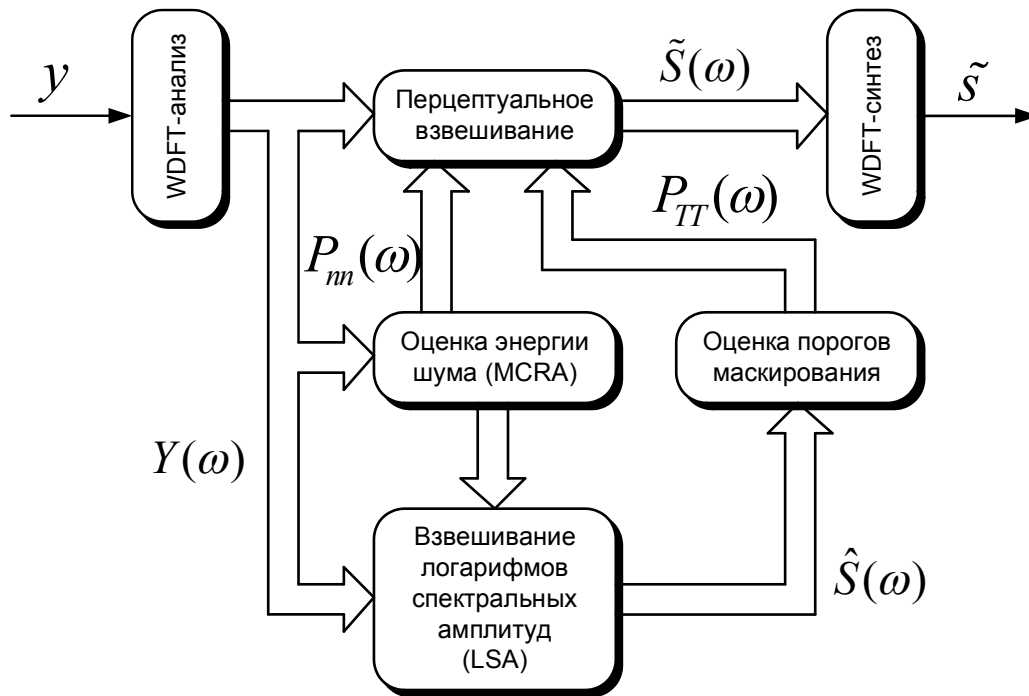
Рис. 11.17 – Уточненная структура комбинированной системы редактирования шума и кодирования широкополосного речевого сигнала

Уточненная структурная схема препроцессора повышения качества речи на базе WDFT представлена на рис. 11.18.

Алгоритм основан на хорошо известном правиле спектрального взвешивания [22-24]:

$$H^{JND}(\omega) = \min \left\{ 1, \sqrt{\frac{P_{TT}(\omega)}{P_{nn}(\omega)} + \xi_{resnoise}} \right\}, \quad (11.11)$$

где JND означает неслышимое шумовое искажение.



y – входной сигнал; $Y(\omega)$ – СПМ входного зашумленного сигнала; $P_{nn}(\omega)$ – СПМ шума; $P_{TT}(\omega)$ – оценка энергии порога маскирования; $\tilde{S}(\omega)$ – СПМ очищенного сигнала; $\hat{S}(\omega)$ – взвешенная оценка СПМ зашумленного сигнала; \tilde{s} – выходной очищенный сигнал.

Рис. 11.18 – Структура математической модели перцептуальной системы подавления шума на базе WDFT [25]

Легко заметить, что, если энергия остаточного шума лежит ниже порога маскирования, выражение под корнем больше единицы и речевой сигнал не искажается так как $H^{JND}(\omega) = 1$. В противном случае, окружающий шум оптимально ослабляется до невоспринимаемого уровня. Для расчета взвешивающих коэффициентов необходима оценка порога маскирования и СПМ шума.

11.5 Математическая модель комбинированной системы компрессора широкополосных речевых сигналов с препроцессором редактирования шума

Формализуем работу системы редактирования шума набором операторов, отвечающих за выполнение алгоритмов препроцессора. С учетом того, что в системе используется еще и обратное частотно-временное преобразование для синтеза очищенного речевого сигнала, то к модели добавится соответствующий оператор (обозначим его D^{-1}).

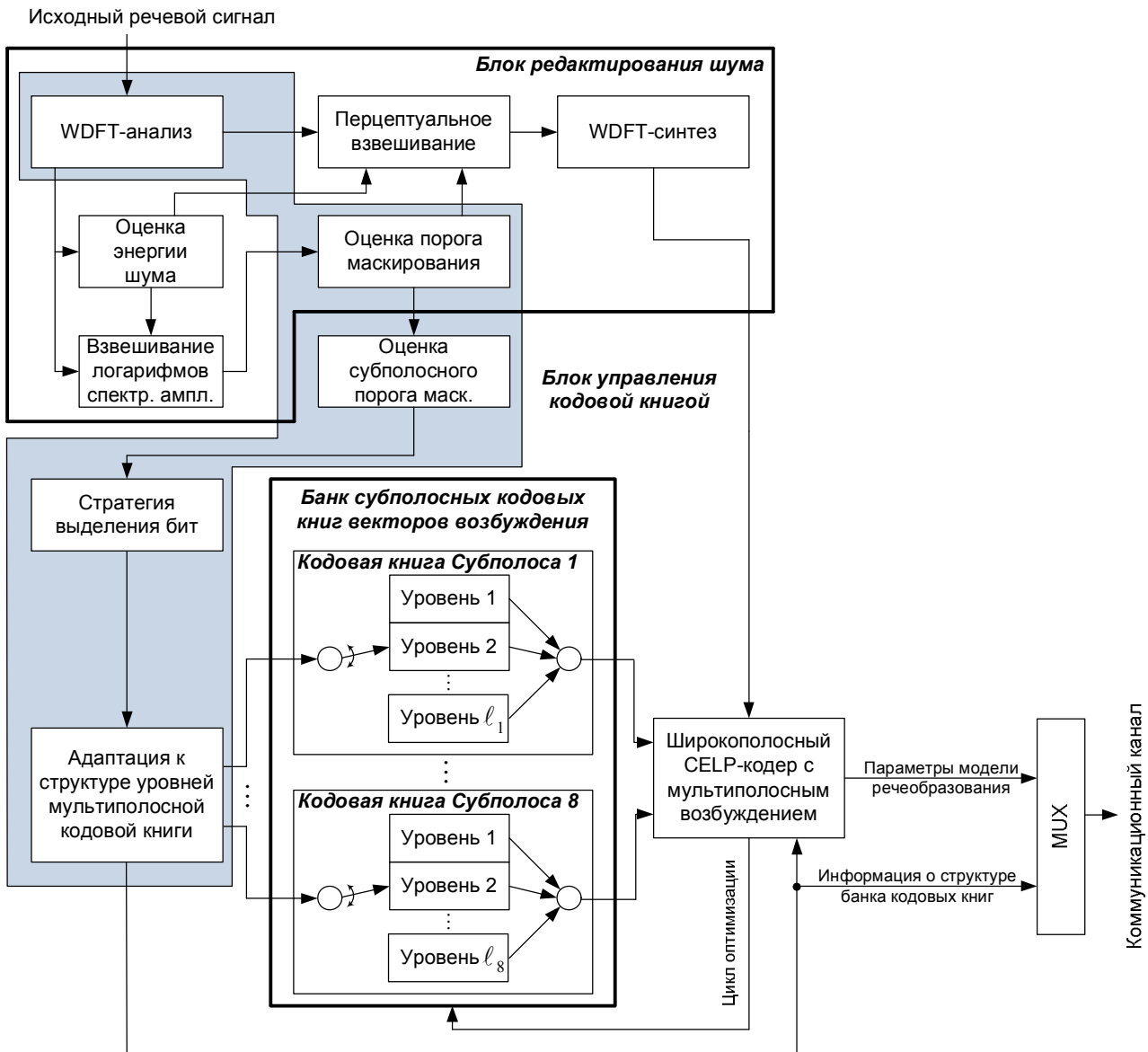


Рис. 11.19 – Комбинированная система кодирования широкополосного речевого сигнала [26,27]

Кроме этого, добавляется правило оценки и взвешивания логарифмов спектральных амплитуд $LSA[\cdot]$ и функция перцептуального взвешивания $H^{JND}[\cdot]$. Расчет порогов маскирования и алгоритм реконфигурации банка кодовых книг остаются прежними. Тогда обобщенный оператор системы редактирования шума можно представить в виде совокупного множества всех вышеперечисленных операций как $NRS[\cdot] = \{D, LSA, MT, H^{JND}, PE, D^{-1}\}$. С учетом дополнительных операторов окончательный вариант математической модели комбинированной системы кодирования широкополосного речевого сигнала может быть записан следующим образом:

$$\{\tilde{s}(n), B, SPE, P_{TT}\} = \{NRS[y(n)]\},$$

$$Bands \in descend \left[\underset{SPE > 0}{B} \right],$$

$$\tilde{s}_w(n) = LP[s(n)], n = 1, 2, \dots, N, 1 \leq \lambda \leq \lambda_c,$$

$$e_{w_0}^{(\lambda)} = \tilde{s}_w(k), l = 1, 2, \dots, L, k = (\lambda - 1) \cdot L + 1, (\lambda - 1) \cdot L + 2, \dots, \lambda \cdot L,$$

$$b = \overline{1, C(Bands)},$$

$$\{g_{b,opt}^{(\lambda)}, i_{b,opt}^{(\lambda)}\} = \arg \min_{g_b^{(\lambda)}, i_b^{(\lambda)}} e_{w_b}^{(\lambda)}(l) = \frac{e_{w_{(b-1)}}^{(\lambda)}(l) - G_b(g_b^{(\lambda)}) \cdot \hat{s}_{w_b}(i_b^{(\lambda)}, depth(b))}{P_{TT,b}}, \quad (11.12)$$

$$e_w^{(\lambda)}(l) = e_{w_0}^{(\lambda)}(l) - \sum_{b=1}^{C(B)} G_b(g_{b,opt}^{(\lambda)}) \cdot \hat{s}_{w_b}(i_{b,opt}^{(\lambda)}),$$

$$e_w(n) = \{e_w^{(1)}(l), e_w^{(2)}(l), \dots, e_w^{(\lambda_c)}(l)\},$$

где $y(n)$ – фрейм зашумленного входного сигнала; $\tilde{s}(n)$ – фрейм очищенного сигнала.

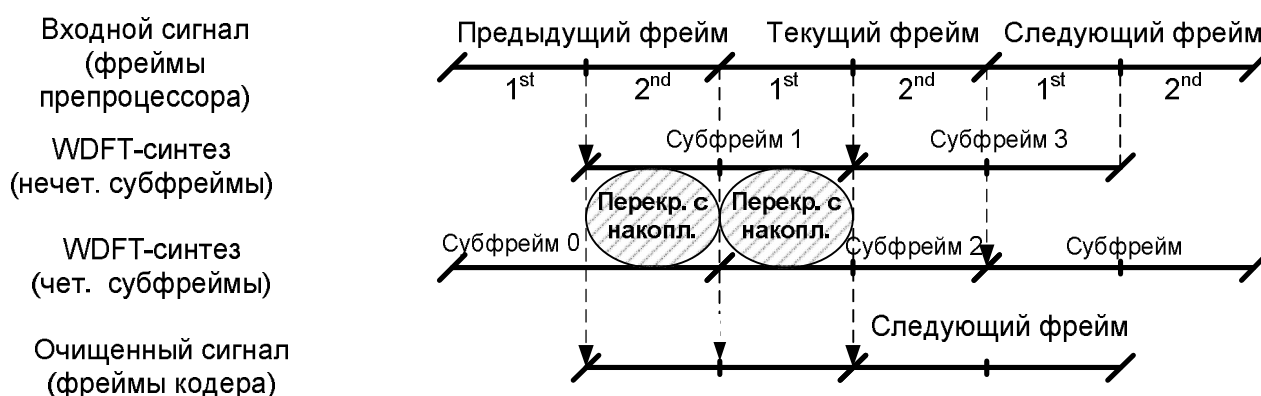


Рис. 11.20 – Формирование фреймов для WDFT-анализа-синтеза

Табл. 11.2 – Характеристика квантования параметров модели речеобразования

Параметр модели	Количество параметров/фрейм	Количество, бит/параметр	Количество, бит/фрейм	Количество, фреймов/с	Скорость потока, бит/с
LSF	16	2	32	50	1600
G_{model}	1	7	7		350
$Delays_{LTP}$	4	7	28		$VAR(0-400)$
$Gains_{LTP}$	4	5	20		$VAR(0-1000)$
PPF	1	4	4		200
$depth$	1 вектор из 8 элементов	1.25	9		450
$Excitation Gains(EG)$	$EG = 4 \cdot \sum_{b \in Bands} 5$	5	$VAR(0-160)$		$VAR(0-8000)$
$Book Indexes (BI)$	$BI = 4 \cdot \sum_{b \in Bands} \log_2(depth(b))$	$VAR(4-8)$	$VAR(0-224)$	$VAR(0-11200)$	
Результирующий поток данных:					$VAR(max. 24200)$

Комбинированная система кодирования широкополосной речевой информации с встроенным препроцессором повышения качества речи может быть представлена в виде структурной схемы, изображенной на рис. 11.19.

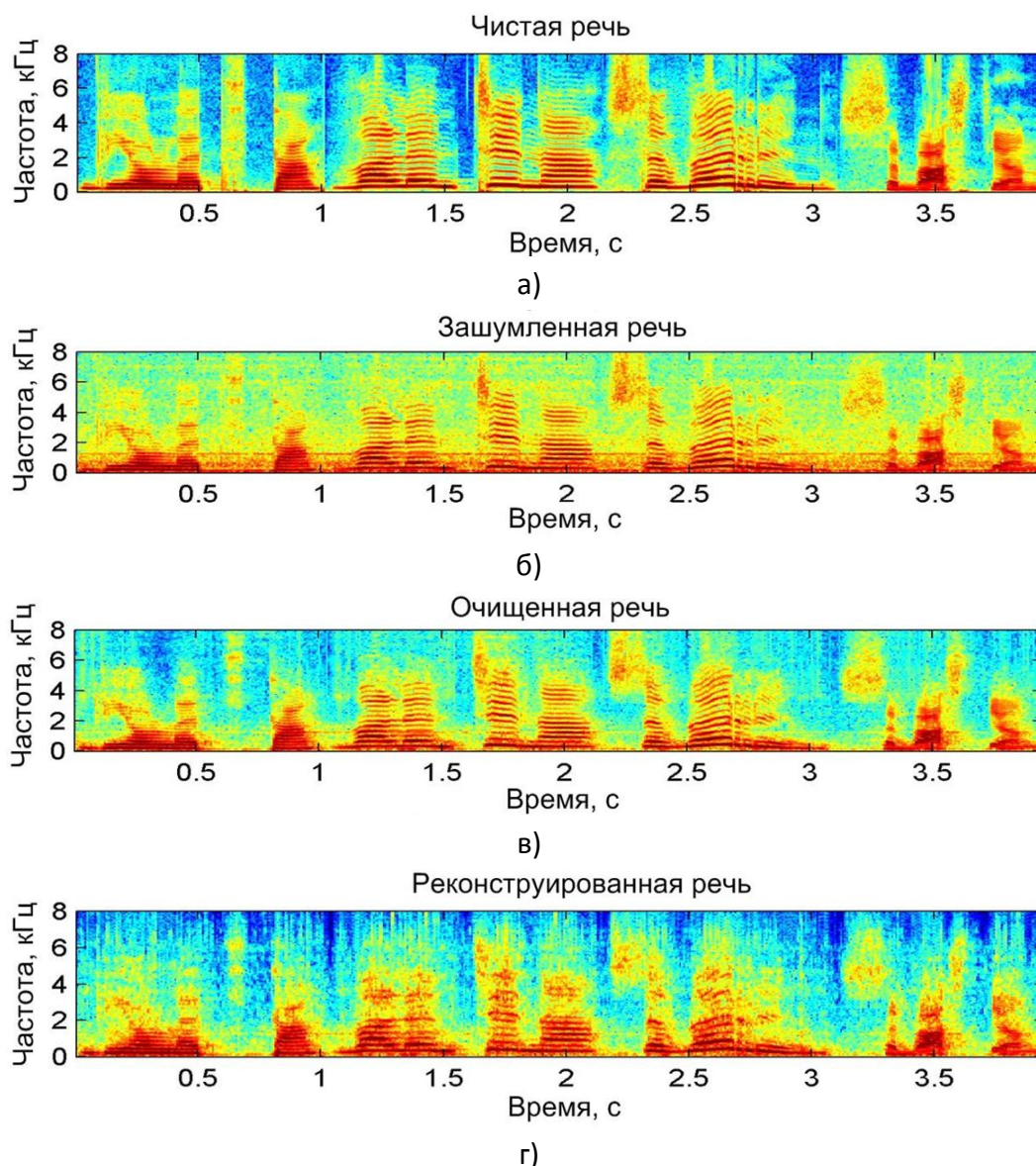


Рис. 11.21 – Сравнение спектрограмм для аддитивного шума типа «шум автомобиля» с уровнем $SNR=5$ дБ: а) чистая речь, б) зашумленная речь, в) очищенная речь, г) реконструированная речь

Препроцессор редактирования шума использует обработку фрейма входного сигнала с длительностью равной длительности фрейма широкополосного кодера. При этом анализ зашумленного речевого сигнала осуществляется фреймами с 50% перекрытием. В блоке WDFT-синтеза осуществляется формирование обработанного фрейма сигнала путем перекрытия с накоплением. Таким образом, система редактирования шума обрабатывает два субфрейма на один фрейм кодера. Для наглядности, формирование входных фреймов системы редактирования шума и фреймов кодера представлено на рис. 11.20.

Как видно из рис. 11.20, блок редактирования шума добавляет алгоритмическую задержку, равную половине фрейма кодера (10 мс), что приводит к общей алгоритмической задержке комбинированной системы кодирования речевой информации, равной 30 мс.

Схема квантования параметров модели речеобразования преобразуется к виду, представленному в табл. 11.2.

В качестве примера на рис. 11.21 представлен сравнительный анализ спектрограмм оригинального речевого фрагмента, зашумленного, очищенного (системой подавления шума) и восстановленного после кодирования комбинированной системой.

11.6 Список использованных источников

1. Makur A., Derivation of Subband Coding Gain: The Most General Case [Electronic resource], Mode of access: www.ntu.edu.sg/home/eamakur/codinggain.pdf, Date of access: 10.05.2007.
2. Zwicker E., Fastl H., Psychoacoustics: facts and models, Berlin: Springer - Verlag, 1990, 354 p.
3. Алдошина И., Основы психоакустики. Субъективные и объективные методы оценки разборчивости речи [Электронный ресурс], Журнал «Звукорежиссер», 2002, Режим доступа: http://rus.625-net.ru/audioproducer/2002/05/theory_3.htm, Дата доступа: 25.05.2006.
4. Kondo A.M., Digital speech: coding for low bit rate communication systems, New York: John Wiley & Sons Inc., 1996, 442 p.
5. Franz S., Mitra S.K., Doblinger G., Frequency estimation using warped discrete Fourier transform, Signal Processing, 2003, vol. 83, pp. 1661–1671.
6. Makur A., Mitra S.K., Warped Discrete-Fourier Transform: Theory and Applications, IEEE Trans. Circuits Systems I., 2001, Vol. 48, No. 9, pp. 1086–1093.
7. Schroeter T., Frequency Warping with Arbitrary Allpass Maps, IEEE Signal Processing Letters, May, 1999, Vol. 6, pp. 116–118.
8. Johnston J.D., Transform coding of audio signals using perceptual noise criteria, IEEE Transactions on Selected Areas Communication, February, 1988, Vol. 6, pp. 314–323.
9. Thiede T. and et al., PEAQ–The ITU standard for objective measurement of perceived audio quality, Journal of AudioEngineering Society, Jan. / Feb., 2000, Vol. 48, pp. 3–29.
10. Bagchi S., Mitra S.K., The Nonuniform Discrete Fourier Transform and Its Applications in Filter Design // IEEE Trans. Circuits Systems II. – 1996. – vol. 43, no. 6. – P. 422–433.
11. Parfieniuk M., Petrovsky A., Warped DFT as the basis for psychoacoustical model, Acoustics, speech and signal processing: proceedings of IEEE int. conf. (ICASSP-2004), Montreal, Canada, May 17–21, 2004, Vol. 4, pp. 185–188.
12. Лившиц М.З., Парфенюк М., Петровский А.А., Широкополосный CELP-кодер с мультиполосным возбуждением и многоуровневым векторным квантованием по кодовой книге с реконфигурируемой архитектурой, Цифровая обработка сигналов, №2, Москва, 2005, с. 20-35.

-
13. Livshitz M.Z., Petrovsky A.A., Perceptually Constrained Variable Bitrate Wideband Speech Coder, The Proc. of EUROCON, Serbia & Montenegro, Belgrade, November 22-24, 2005, pp.1296-1299.
 14. Livshitz M., Petrovsky A., Synthesis of Codebooks with Perceptually Monitored Structure for Multiband CELP-Coders, in "Biometrics, Computer Security Systems and Artificial Intelligence Applications" // edited by Saeed K., Pejas J., Mosdorf R., Springer Science + Business Media, New York, USA, 2006, pp.57-67.
 15. Petrovsky A., Parfieniuk M., Bielawski K., Psychoacoustically Motivated Non-uniform Cosine Modulated Polyphase Filter Bank, 2nd International Workshop on Spectral Methods and Multi-rate Signal Processing (SMMSP 2002), Toulouse, France, September 7-8, 2002, pp. 95-101.
 16. Лившиц М.З., Векторное квантование в широкополосных кодерах речи, Известия Белорусской инженерной академии, №1(15)/2, 2003, с.281-284.
 17. Huffman D.A., Method for the construction of minimum redundancy codes, The Proceedings of IRE., September, 1952, Vol. 40, pp. 1098-1101.
 18. Bernard A., Alwan A., Perceptually Based and Embedded Wideband CELP Coding of Speech, Proceedings of Eurospeech, Budapest, Hungary, 1999, pp. 1543-1546.
 19. Boll S.F., Suppression of acoustic noise in speech using spectral subtraction, IEEE Trans. Acoust., Speech, Signal Processing, 1979, Vol. 27, No. 2, pp. 113-120.
 20. Evans N.W.D., Mason J.S.D., Liu W.M., Fauve B., An Assessment on the Fundamental Limitations of Spectral Subtraction, Proceedings of the 9th International Conference on Spoken Language Processing (Interspeech 2006 - ICSLP), Pittsburgh, Pennsylvania, USA, September 17-21, 2006, Vol. I, pp. 145-148.
 21. Virag N., Single channel speech enhancement based on masking properties of the human auditory system, IEEE Transactions on Speech and Audio Processing, 1999, Vol. 7, No. 2, pp. 126-137.
 22. Gustafsson S., Jax P., Kamphausen A., Vary P., A postfilter for echo and noise reduction avoiding the problem of musical tones, IEEE Int. Conf. on Acoustic, Speech, and Signal Processing ICASSP'99, Phoenix, USA, 1999, Vol. 2, pp. 873-876.
 23. Gustafsson S., Jax P., Vary P., A novel psychoacoustically motivated audio enhancement algorithm preserving background noise characteristic, IEEE Int. Conf. on Acoustic, Speech, and Signal Processing ICASSP'98, Seattle, USA, 1998, Vol. 1, pp. 397-400.
 24. Gustafsson S., Martin R., Jax P., Vary P., A psychoacoustic approach to combined acoustic echo cancellation and noise reduction, IEEE Transactions on Speech and Audio Processing, 2002, Vol. 10, No. 5, pp. 245-256.
 25. Petrovsky A.A., Parfieniuk M., Borowicz A., Warped DFT based perceptual noise reduction system, Signal processing: proceedings of 116th AES convention, Germany, Berlin, May 8-11, 2004, Preprint № 6035, 16 pages.

-
26. Livshitz M.Z., Petrovsky A.A., An Overcomplete WDFT-based Perceptually Constrained Variable Bit Rate Wideband Speech Coder with Embedded Noise Reduction System, The Proc. of XI International Conference "SPEECH AND COMPUTER", St. Petersburg, Russia, June 25-29, 2006, pp.343-348.
 27. Livshitz M.Z., Parfieniuk M., Petrovsky A.A., Real-Time Implementation of Combined Noise Reduction and Wideband Speech Coding Systems based on the DSP TMS320C6713, in "New Trends in Audio and Video" // edited by Andrzej Dobrucki, Alexander Petrovsky, Wladislaw Skarbek, Politechnika Bialostocka, Bialystok, 2006, vol. I, pp.473-478.

12 ГИБРИДНОЕ КОДИРОВАНИЕ РЕЧЕВЫХ СИГНАЛОВ НА ОСНОВЕ ДПФ СИНХРОНИЗИРОВАННОГО С ИЗМЕНЕ- НИЕМ ЧАСТОТЫ ОСНОВНОГО ТОНА

Павловец А.Н., Петровский А.А.

АННОТАЦИЯ

Описывается метод кодирования речевых сигналов на базе гибридных техник. Данный подход обусловлен тем фактом, что методы обработки речи в частотной и временной областях обладают своими преимуществами и недостатками. Рассматриваемая здесь гибридная модель использует преимущества декомпозиции речи на периодическую и аperiodическую компоненты для участков вокализованной речи и CELP-модели – для нестационарных (переходных) участков. Анализ речевого сигнала осуществляется с помощью дискретного преобразования Фурье, согласованного с изменением контура частоты основного тона.

Ключевые слова – гибридная модель речевого сигнала, декомпозиция на периодическую и аperiodическую компоненты, компрессия речи, гармонический анализ.

СОДЕРЖАНИЕ

12.1 Введение.....	261
12.2 Гибридная модель речевого сигнала.....	262
12.3 Определение параметров модели.....	264
12.3.1 Гармонический анализ.....	264
12.3.2 Преобразование Фурье, согласованное с изменением контура частоты основного тона.....	266
12.3.3 Повышение точности ДПФ, согласованного с изменением контура частоты основного тона	267
12.3.4 Метод определения частоты основного тона с одновременной декомпозицией речевого сигнала на периодическую и аperiodическую компоненты	270
12.3.5 Анализ переходных сегментов	275
12.4 Синтез речи по параметрам модели “гармоники + шум”	275
12.5 Гибридный вокодер со скоростью передачи 2,4/4,8 кбит/с, основанный на декомпозиции речевого сигнала на периодическую и аperiodическую компоненты.....	276
12.6 Список использованных источников.....	282

12.1 Введение

Кодеры формы сигнала и схемы, основанные на линейном предсказании, используют временное представление сигнала и обеспечивают хорошее качество при скоростях передачи 6 кбит/с и выше. В диапазоне скоростей ниже 4 кбит/с, который требуется для беспроводных и

спутниковых систем коммуникаций, при сохранении разборчивости у таких кодеров отмечается существенная деградация качества синтезируемой речи. Многочисленные исследования показали, что при низких скоростях передачи более эффективным является параметрическое кодирование, при котором модель речевого сигнала учитывает компактный набор параметров, особенно значимых для восприятия [1]. Наибольшее распространение получили кодеры, использующие в качестве параметров различные характеристики представления сигнала в частотной области – спектральные кодеры [2]. Исследования в этой области сосредоточились в основном на группе синусоидальных кодеров, родоначальниками в которой считаются Harmonic Coder [3], Sinusoidal Transform Coder [4] и Multiband Excitation Coder [5].

В первых спектральных кодерах сегменты речевого сигнала классифицировались либо как вокализованные, либо как шумовые. Такое допущение позволяло довольно просто реализовать анализ-синтез, но существенно ограничивало качество выходного сигнала. В гармоническо-шумовых кодерах использовалась методика разбиения спектра на полосы с независимой классификацией вокализованности [5]. Каждый фрейм сигнала при синтезе комбинировался из двух компонент: гармонической и шумовой. Эта же идея реализована в ряде других схем кодирования речи [6]. Дальнейшее совершенствование привело к новому усложнению классификации фреймов [2]. Параллельно стал развиваться предложенный в [7] подход, в соответствии с которым вместо классификации сигнал разделялся на детерминистскую и шумовую компоненты. Детерминистская составляющая представляет собой гармонический спектр, а шумовая – сигнал-остаток после ее выделения [8,9]. Данный подход характеризуется такими преимуществами, как универсальность, высокое качество выходного сигнала, возможность эффективной реализации. В [10] показано еще одно важное достоинство, которое в полной мере не реализуется ни в одном из известных в настоящее время параметрических кодеров – возможность учета закономерностей психоакустики.

Несмотря на то, что, создавая низкоскоростные устройства компрессии речи на базе гармонической и шумовой моделей, можно добиться хорошей разборчивости реконструированной речи, данные модели несостоятельны при анализе речевых отрезков, содержащих взрывные звуки, наличии в сегменте одновременно гласных и глухих согласных (переходные сегменты), присутствии каких-то локальных явлений. Таким образом, вопрос получения реконструированной речи высокого качества при низкой скорости передачи всё ещё актуален.

12.2 Гибридная модель речевого сигнала

Большинство сигналов в природе, включая речь и музыку, могут быть описаны при помощи гармонической модели, которая определяется следующим набором параметров: фундаментальной частотой, амплитудой и фазой каждой частотной компоненты. Гармонический сигнал генерируется серией синусоид или гармонических компонент, частоты которых являются целочисленным кратным некоторой фундаментальной частоты. Данная модель является весьма эффективным решением для большого количества приложений кодирования сигнала, так как позволяет представить сигнал с помощью достаточно компактного набора параметров – частоты основного тона, амплитуд и фаз гармоник:

$$h(i) = \sum_{k=1}^K A_k \cos\left(\frac{2\pi k i}{F_s} F_0 + \theta_k\right), \quad (12.1)$$

где F_0 – частота основного тона (ЧОТ), A_k – амплитуда k -ой гармоники основного тона, θ_k – фаза k -ой гармоники основного тона, K – количество гармоник, F_s – частота дискретизации.

Первые попытки представления речевого сигнала с помощью гармонической модели датируются началом 80-х годов [3]. В дальнейшем, в системах анализа-синтеза речи данное представление стало уточняться и дополняться описанием сигнала-остатка в форме шумовой модели [11], что позволяет повысить точность представления речевого сигнала, а вместе с этим и качество:

$$s(i) = h(i) + r(i). \quad (12.2)$$

Некоторые сегменты речевого сигнала сложно разделить на периодическую и аperiodическую составляющие, используя гармоническую и шумовую модели. Это происходит в том случае, если анализируемый речевой отрезок носит переходный характер (рис. 12.1).

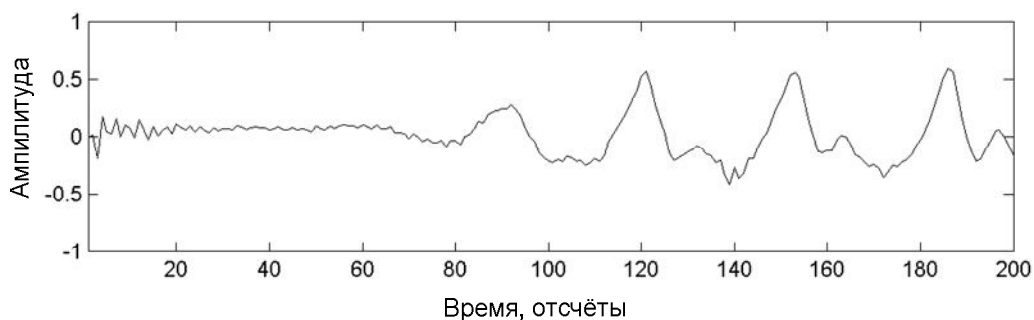


Рис. 12.1 – Переходный сегмент речевого сигнала

Таким образом, следующей ступенью развития представления речевого сигнала стала гибридная модель [12], предусматривающая три возможных класса для сегмента речи – вокализованный, невокализованный, переходный. Особенностью её является анализ-синтез переходных сегментов во временной области, в то время как вокализованные и невокализованные сегменты обрабатываются в частотной области.

Модель обработки речевого сигнала, рассматриваемая в данном разделе, предполагает классификацию речевого сегмента с точки зрения возможности декомпозиции его на гармоническую и шумовую компоненты. Признаком такой возможности является вокализованность сегмента (рис. 12.2).

В такой модели важнейшим аспектом является корректное определение гармонической и шумовой компонент. Существуют различные подходы к их разделению. Так, например, в [11] сегмент речи представлен суммой гармонической и шумовой составляющих, спектры которых находятся соответственно до и после так называемой максимальной частоты вокализованности. Подход [5] характеризуется определением вокализованности в определённых частотных полосах.

Данные подходы не вполне адекватно описывают речевой сигнал, поскольку в них используются бинарные решения о вокализованности сигнала в целой полосе частот. Более точно

анализ реализован в [13,14], где периодическая и аperiodическая составляющие речевого сигнала разделяются посредством итеративного алгоритма, использующего экстраполяцию спектра. В [15] для декомпозиции сигнала применяется гармонический фильтр, использующий принцип максимального правдоподобия. Особенностью этого фильтра является использование окон переменной длины, согласованных с периодом основного тона. Методы [13-15] показали хорошие результаты при анализе синтетической речи.

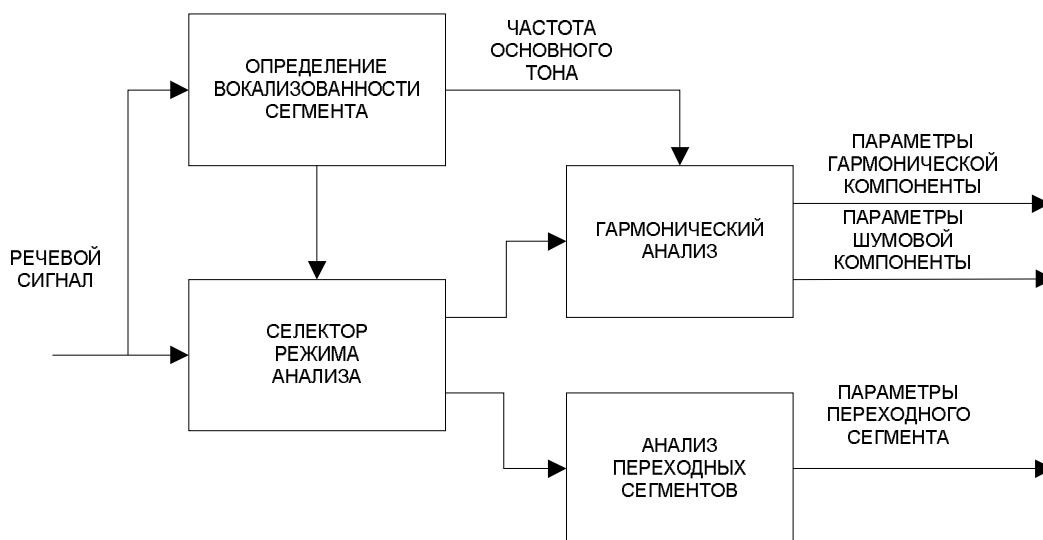


Рис. 12.2 – Модель обработки речевого сигнала

Отличие рассматриваемой модели анализа-синтеза речи от вышеупомянутых состоит в использовании для декомпозиции речевого сигнала дискретного преобразования Фурье, согласованного с изменением контура частоты основного тона [16-18]. Этот подход позволяет разделить гармоническую и шумовую составляющие во всём речевом спектре.

Методика определения параметров модели будет рассмотрена ниже.

12.3 Определение параметров модели

12.3.1 Гармонический анализ

Анализ истории создания устройств параметрической обработки речи показывает, что наилучшее качество синтезированного речевого сигнала обеспечивают те из них, в которых анализ-синтез речевого сегмента осуществляется согласно модели, более полно отражающей природу данного сегмента. Так, гласные и озвученные согласные звуки довольно точно можно описать с помощью гармонической модели, шипящие – с помощью шумовой. Несмотря на значительные успехи на данном направлении [1], окончательного решения проблема так и не получила. В первую очередь это связано с тем, что в большинстве существующих подходов сегмент речи полагается стационарным, либо классифицируется как полностью вокализованный (невокализованный), что далеко не всегда верно. Так, отрезок речи, богатой эмоциями, далеко не стационарен как по отношению к частоте основного тона, так и по отношению к амплитудам гармоник. Данный факт иллюстрирует рис. 12.3, на котором показаны исходный и реконструированный согласно формуле (12.1) речевой сегмент.

Определение амплитуд и фаз гармоник в данном случае велось с помощью анализа в области частот гармоник основного тона:

$$X(k) = \sum_{n=0}^{N-1} x(n)w(n)e^{-j\frac{2\pi nkF_0}{F_s}}, k = 1 \dots K, \quad (12.3)$$

где $X(k)$ – спектральный отсчет k -ой гармоники (коэффициент Фурье), F_s – частота дискретизации, $w(n)$ – временное окно.

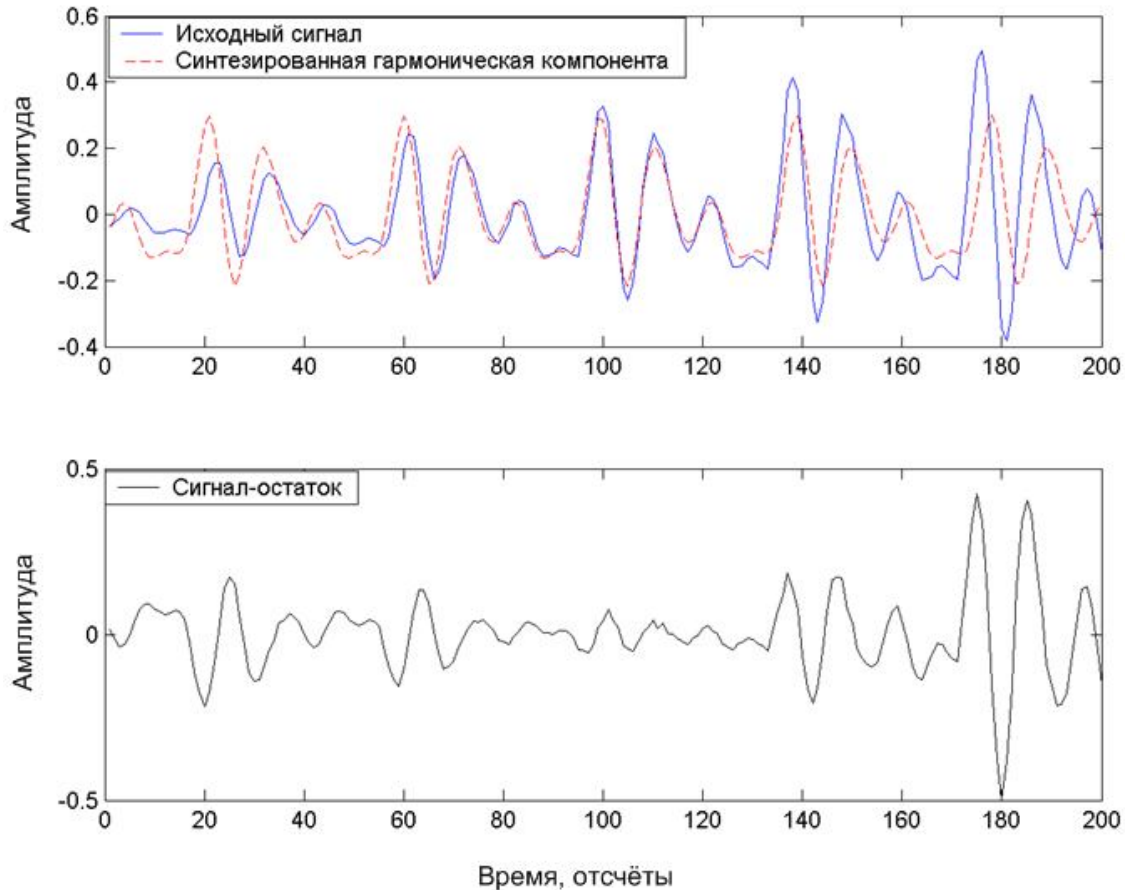


Рис. 12.3 – Иллюстрация недостатков стационарной модели

Очевидно, что в рамках традиционной гармонической модели вследствие неточного анализа в шумовую компоненту попадает значительная часть периодической компоненты, что приводит к последующим искажениям при реконструкции сигнала.

Таким образом, можно утверждать, что в системах обработки речевой информации на основе отдельной обработки компонент различной природы для улучшения качества декомпозиции, необходимо учитывать нестационарность речевого сигнала.

При достаточно малой величине окна анализа (до 25 мс) можно предположить, что изменение частоты основного тона носит линейный характер:

$$h(i) = \sum_{k=1}^K A_k \cos\left(\frac{2\pi ki}{F_s}\left(F_0 + \frac{\Delta F_0 i}{2N}\right) + \theta_k\right), \quad (12.4)$$

где ΔF_0 – изменение частоты основного тона за N отсчётов.

Следующим уточнением гармонической модели речи будет введение в формулу (12.4) фактора, учитывающего изменение (нарастание либо затухание) гармонических амплитуд с течением времени [19]:

$$h(i) = \sum_{k=1}^K A_k e^{-\beta i} \cos\left(\frac{2\pi k i}{F_s} \left(F_0 + \frac{\Delta F_0 i}{2N}\right) + \theta_k\right), \quad (12.5)$$

где β – фактор изменения; предполагается, что значения гармонических амплитуд эволюционируют по экспоненциальному закону.

Величину фактора изменения амплитуд β можно определить следующим образом [19]:

$$\beta = \frac{\ln(x)}{T_0}, \quad x = \frac{S_0^T S_{T_0}}{S_{T_0}^T S_{T_0}}, \quad (12.6)$$

где $S_0 = [|s_0|, \dots, |s_{N-T_0-1}|]^T$, $S_{T_0} = [|s_{T_0}|, \dots, |s_{N-1}|]^T$, T_0 – значение периода основного тона в отсчётах, s_0, s_1, \dots, s_{N-1} – отсчёты речевого сигнала, N – длина сегмента.

Таким образом, данную модель периодического речевого сигнала полностью характеризует набор следующих параметров: частота основного тона F_0 , отклонение частоты основного тона ΔF_0 , амплитуды A и фазы θ гармоник основного тона, фактор изменения амплитуд β .

12.3.2 Преобразование Фурье, согласованное с изменением контура частоты основного тона

С целью более точного представления периодической компоненты речи предлагается заменить классический спектральный анализ на основе кратковременного преобразования Фурье спектральным анализом, согласованным с изменением контура частоты основного тона (следящим анализом) [17].

В случае следящего анализа перестраивается ядро преобразования (12.3). Так, аргумент под экспонентой может быть записан следующим образом:

$$\varphi(n, k) = \frac{2\pi n k F_0}{F_s}, \quad (12.7)$$

Если частота основного тона меняется линейно на анализируемом фрейме на величину ΔF_0 , тогда аргумент под экспонентой будет вычисляться по формуле:

$$\varphi(n, k) = \frac{2\pi n k F_0}{F_s} + \frac{2\pi k \Delta F_0 n^2}{2N F_s} = \frac{2\pi n k}{F_s} \left(F_0 + \frac{\Delta F_0 n}{2N}\right), \quad (12.8)$$

Таким образом, следящий гармонический анализ задается формулой:

$$X(k) = \sum_{n=0}^{N-1} x(n) w(n) e^{-j\varphi(n, k)}, \quad k = 1 \dots K, \quad (12.9)$$

Амплитуды и начальные фазы гармоник вычисляются в соответствии со следующими выражениями:

$$A_k = \sqrt{\operatorname{Re}^2 X(k) + \operatorname{Im}^2 X(k)}, \quad \Phi_k = -\arctan\left(\frac{\operatorname{Im} X(k)}{\operatorname{Re} X(k)}\right). \quad (12.10)$$

Пример 12.1. Следящий анализ сегмента синтетического звука "А" с частотой основного тона $F_0 = 300$ Гц.

```
Fs = 8000;           %частота дискретизации
N = 200;            %длина сегмента
F0 = 300;           %частота основного тона в начале сегмента
F01 = 315;          %частота основного тона в конце сегмента
amps = [0.1205,0.1065,0.2427,0.1759,0.0733,0.0134,0.0002,0.001,0.001,...
        ...0.001,0.001,0.002]; %распределение амплитуд гармоник основного тона
        %соответствует звуку "А"
phases = [1.7509,0.0498,2.9331,-1.5811,1.5071,-3.0232,0.437,1.4083,...
          ...-1.6895,2.6679,-0.2464,-2.471];%распределение начальных фаз
          %гармоник основного тона соответствует звуку "А"
deltaF0 = F01-F0;
Frame = zeros(1,N);
K = min([floor(Fs/2/F0) floor(Fs/2/(F0+deltaF0))]); %количество гармоник
        %основного тона
amps_est = zeros(1,K); %резервирование памяти под массив
        %рассчитанных амплитуд гармоник
phases_est = zeros(1,K);%резервирование памяти под массив рассчитанных фаз
        %гармоник
a1 = 2*pi*F0/Fs;
a2 = 2*pi*deltaF0/(Fs*N*2);
i = 0:N-1;
for k = 1:K,          %синтез сегмента звука "А" для последующего анализа
    Frame = Frame+amps(k).*cos((a2.*(i.^2)+a1.*i).*k+phases(k));
end;
w = hamming(N);
w = w./(sum(w)/2);
Frame_w = Frame.*w'; %анализируемый сегмент взвешивается окном Хэмминга
for k = 1:K,
    V = Frame_w.*exp(-j*((a2.*(i.^2)+a1.*i)*k)); %выполняется ДПФ,
        %согласованное с контуром частоты основного тона
    amps_est(k) = abs(sum(V)); %рассчитываются амплитуды гармоник
        %основного тона
    phases_est(k) = angle(sum(V)); %рассчитываются фазы гармоник
        %основного тона
end;
```

Главное преимущество анализа на основе ДПФ, согласованного с изменением контура частоты основного тона (Pitch Tracking DFT (PTDFT) [16,20]), в отличие от кратковременного ДПФ (Short-Time Fourier Transform (STFT)) – частота основного тона точно совпадает с частотным отсчетом соответствующего коэффициента преобразования. Спектральный анализ проводится в области гармоник частоты F_0 .

12.3.3 Повышение точности ДПФ, согласованного с изменением контура частоты основного тона

Поскольку частоты синусоидальных компонент в (12.5) изменяются, такая система базовых функций является неортогональной, что приводит к снижению точности анализа. Для уменьшения этого эффекта и одновременного синтеза тональной компоненты в качестве ис-

ходного сигнала при определении параметров k -й гармоники использовалась разность оригинального речевого сигнала и уже синтезированной части тональной компоненты:

$$s_{n,k}(i) = s_n(i) - \sum_{m=1}^{k-1} A_n(m) \cos[(a_{2,n}i^2 + a_{1,n}i)k + \Phi_n(m)], \quad i = \overline{0, N-1} \quad (12.11)$$

где $s_n(i)$ – оригинальный речевой сигнал; $a_{1,n}$, $a_{2,n}$ – параметры, обеспечивающие линейное изменение частоты базовых функций анализа в соответствии с рассчитанной траекторией частоты основного тона:

$$a_{2,n} = \frac{2\pi(F_{0,n} - F_{0,n}^{(1)})}{2F_s N}, \quad a_{1,n} = \frac{2\pi F_{0,n}}{F_s} \quad (12.12)$$

где $F_{0,n}$, $F_{0,n}^{(1)}$ – частота основного тона в конце и в начале n -го окна анализа.

Для первой итерации полагается $s_{n,1}(i) = s_n(i)$. Аналогичный подход для случая несогласованного спектрального анализа описан в [21].

Применение традиционных временных окон в выражении (12.9) не обеспечивает наилучших результатов. Точность определения параметров сигнала может быть повышена, если форма окна будет согласовываться с изменением частоты основного тона. Например, при использовании в качестве прототипа окна Кайзера, согласованное окно описывается следующим образом:

$$w_n(i) = \frac{I_0\left(\beta \sqrt{1 - \left[\frac{2x(i) - N + 1}{N - 1}\right]^2}\right)}{I_0(\beta)}, \quad (12.13)$$

где $I_0(\beta)$ – функция Бесселя нулевого порядка; β – параметр окна;

$$x(i) = \frac{a_{2,n}(N - 1 - i)^2 + a_{1,n}(N - 1 - i)}{a_{2,n}(N - 1) + a_{1,n}}, \quad (12.14)$$

Пример 12.2. Синусоида с линейно изменяющейся частотой и соответствующее ей окно Кайзера (параметр $\beta = 7$):

```
F = 30; %значение частоты
deltaF = 55; %приращение частоты
Fs = 8000; %частота дискретизации
N = 250; %длина сегмента синусоиды
beta = 7; %параметр окна Кайзера
i = 0:N-1;
a1 = 2*pi*F/Fs;
a2 = 2*pi*deltaF/(2*Fs*N);
s = sin(a1.*i+a2.*i.^2); %отсчеты синусоиды с линейно
%меняющейся частотой

bes = abs(besseli(0,beta)); %знаменатель окна
x = (a2.*(N-1-i).^2 + a1.*(N-1-i))./(a2*(N-1)+a1);
xn = [(2*x-N+1)/(N-1)].^2;
w = besseli(0,beta*sqrt(1-xn))/bes; %отсчеты согласованного окна Кайзера
subplot(2,1,1);
```

```

plot(i,s,'k');           %график синусоиды
subplot(2,1,2);
plot(i,w,'k');          %график согласованного окна Кайзера

```

Результат приведен на рис. 12.4. При постоянной частоте основного тона данное окно вырождается в окно Кайзера.

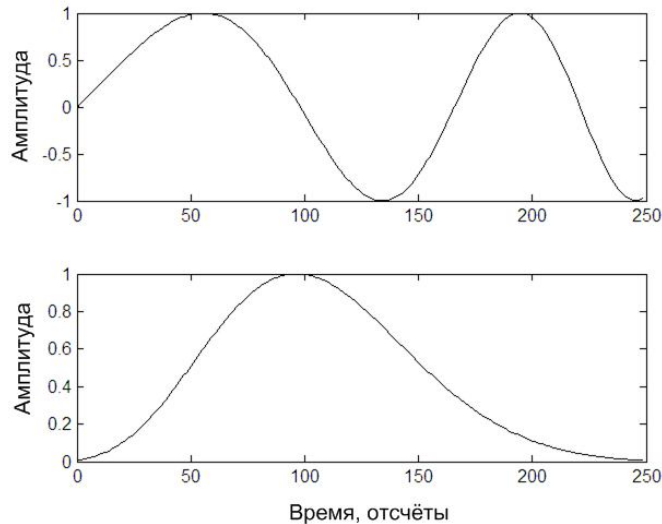


Рис. 12.4 – Синусоида с линейно изменяющейся частотой (вверху) и соответствующая ей форма окна Кайзера (внизу)

Возможно также применение матрицы компенсации неортогональности преобразования [16]. Формирование матрицы и алгоритм коррекции результатов следующий:

- а) вычислить частотные отсчеты $X(m)$;
- б) осуществить синтез косинусоидального сигнала соответствующей гармонике m :

$$H_{n,m}(i) = \cos[(a_{1,n}i + a_{2,n}i^2)m]; \quad (12.15)$$

- в) вычислить влияние m частотного отсчета на другие:

$$e_{lm} = \sum_{i=0}^{N-1} H_{m,n}(i)w(i)[\cos(l(a_{1,n}i + a_{2,n}i^2)) - j\sin(l(a_{1,n}i + a_{2,n}i^2))]; \quad (12.16)$$

- г) повторить пункты б) и в) для всех частотных отсчетов;
- д) построить матрицу компенсации неортогональности преобразования:

$$E_1 = \begin{bmatrix} e_{11} & e_{12} & e_{13} & \cdots & e_{1m} \\ e_{21} & e_{22} & e_{23} & \cdots & e_{2m} \\ e_{31} & e_{32} & e_{33} & \cdots & e_{3m} \\ \vdots & \vdots & \vdots & & \vdots \\ e_{l1} & e_{l2} & e_{l3} & \cdots & e_{lm} \end{bmatrix}; \quad (12.17)$$

- е) повторить пункты б) – д) алгоритма лишь с той разницей, что в пункте б) генерируется синусоидальный сигнал, в пункте д) формируется матрица компенсации E_2 ;
- ж) рассчитать новые значения частотных отсчетов в соответствии с формулой:

$$\tilde{X} = E_1^{-1}\text{Re } X - E_2^{-1}\text{Im } X. \quad (12.18)$$

12.3.4 Метод определения частоты основного тона с одновременной декомпозицией речевого сигнала на периодическую и аperiodическую компоненты

Поскольку гармоническая и шумовая модели кардинально отличаются друг от друга, важным аспектом является точная сепарация гармонической и шумовой компонент речи.

Рассмотрим влияние точности определения частоты основного тона на качество сепарации. Для этого был использован синтетический тестовый полигармонический сигнал со следующими параметрами: $F_0 = 315$ Гц, начальное распределение фаз и амплитуд гармоник основного тона соответствует звуку “А” (рис. 12.5).

При определении значений амплитуд и фаз гармоник тестового сигнала с использованием формул (12.8) – (12.10) вместо правильного значения $F_0 = 315$ Гц подставлялись значения, моделирующие ошибку в определении частоты основного тона, отклонения от верного значения составляли от 0,1 до 10 Гц. Сигнал, синтезированный с использованием рассчитанных значений амплитуд и фаз гармоник, далее вычитался из тестового:

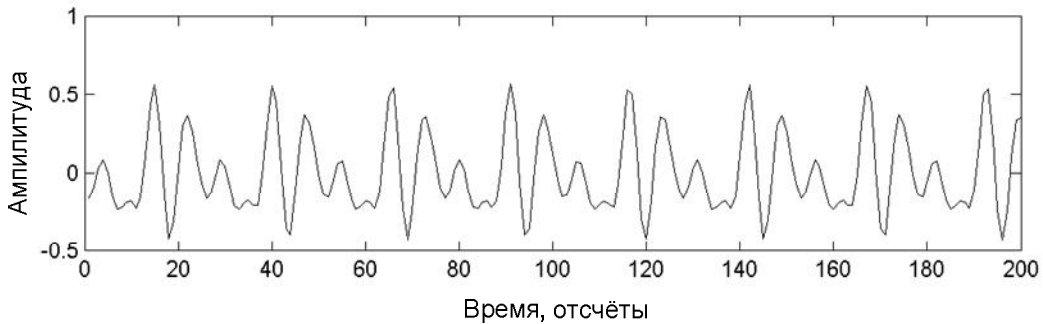


Рис. 12.5 – Сегмент тестового сигнала

$$r(i) = s(i) - \hat{h}(i). \quad (12.19)$$

В качестве показателя точности будет использоваться отношение “гармоники / шум”:

$$HNR = 10 \lg \frac{E_h}{E_n}, \quad (12.20)$$

где E_h и E_n – энергии гармонической и шумовой компоненты соответственно.

Результат данного эксперимента приведен на рис. 12.6. Уже при величине моделируемой ошибки определения частоты основного тона равной 5 Гц величина HNR приближается к значению 2 дБ, что совершенно неудовлетворительно. Таким образом, для качественной обработки речевого сигнала с использованием метода его декомпозиции требуется метод определения частоты основного тона с высокой степенью точности.

Рассмотрим спектрограмму некоторого речевого фрагмента (рис. 12.7). Очевидно, что большая часть данного фрагмента имеет гармоническую структуру с фундаментальной частотой, изменяющейся в области примерно 300 – 400 Гц. Рассмотрим зависимость отношения “гармоники / шум” от предполагаемой частоты основного тона. Для этого выберем один из вокализованных сегментов данного фрагмента речи и проведём анализ по формулам (12.5) – (12.10), (12.19), (12.20) для типичных значений частоты основного тона речи от 50 до 500 Гц с шагом 1 Гц. Результат приведен на рис. 12.8.

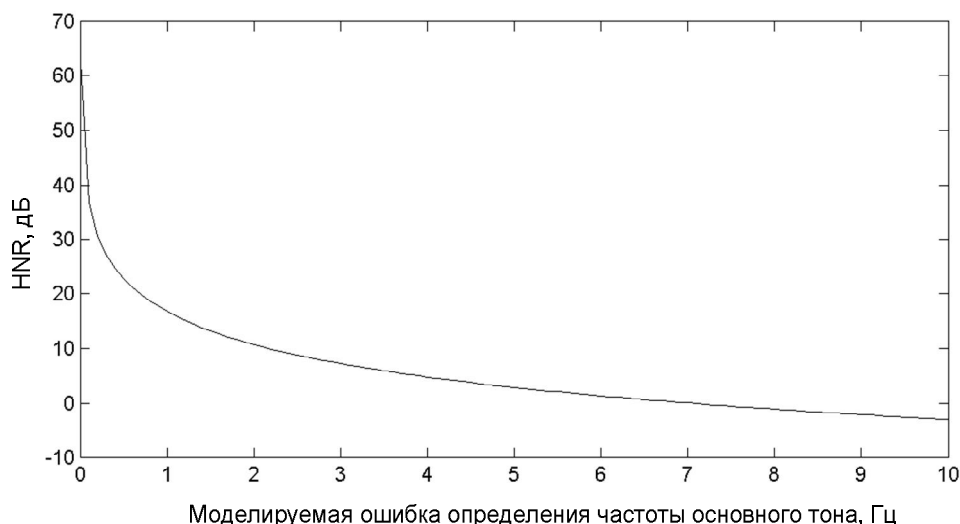


Рис. 12.6 – Зависимость величины отношения “гармоники/шум” от погрешности определения ошибки частоты основного тона

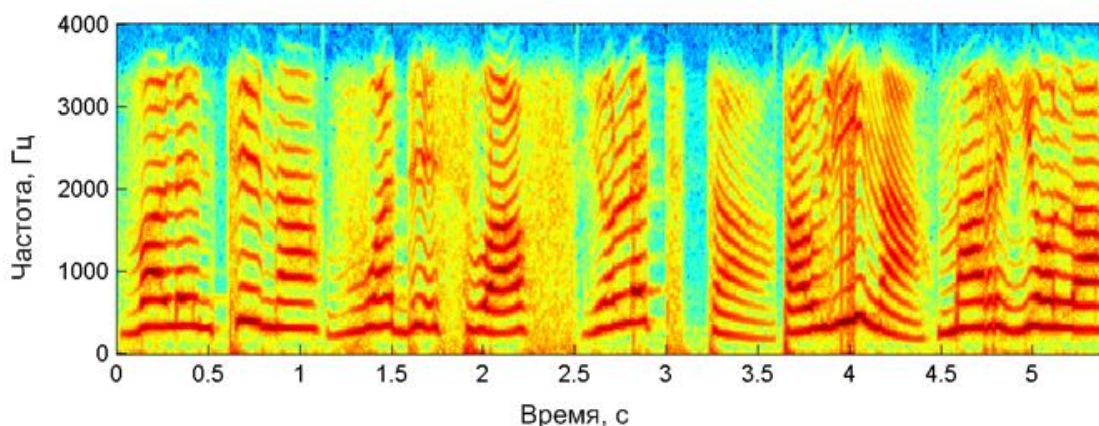


Рис. 12.7 – Спектрограмма речевого фрагмента

Из рис. 12.8 видно, что зависимость $HNR(F_0)$ имеет локальный максимум в точке, равной частоте основного тона данного сегмента речи и носит унимодальный характер в её окрестности (область 280 – 390 Гц). Эксперименты показали, что характер данной зависимости является общим для вокализованных сегментов.

Таким образом, параметры гармонической компоненты будут иметь оптимальные значения при:

$$F_0^{opt} = \arg \max(HNR(F_0)), \quad F_{0l} \leq F_0 \leq F_{0r}, \quad (12.21)$$

где диапазон $[F_{0l}, F_{0r}]$ – некоторая окрестность фундаментальной частоты.

Таким образом целесообразна следующая методика определения параметров гармонической компоненты: сначала проводится приблизительная оценка частоты основного тона и отслеживание её контура, а затем в окрестности этой оценки проводится поиск максимума HNR методом анализа-через-синтез по формулам (12.5) – (12.10), (12.19), (12.20).

Алгоритм определения параметров гармонической компоненты выглядит следующим образом.

Шаг 1. Рассчитать значения нормализованной автокорреляционной функции (НАКФ) для всех допустимых значений периода основного тона:

$$\psi(p) = \frac{\sum_{j=1}^N s_j s_{j+p}}{\sqrt{\sum_{j=1}^N s_j^2 \sum_{j=1}^N s_{j+p}^2}}, \quad (12.22)$$

где p – период тона.

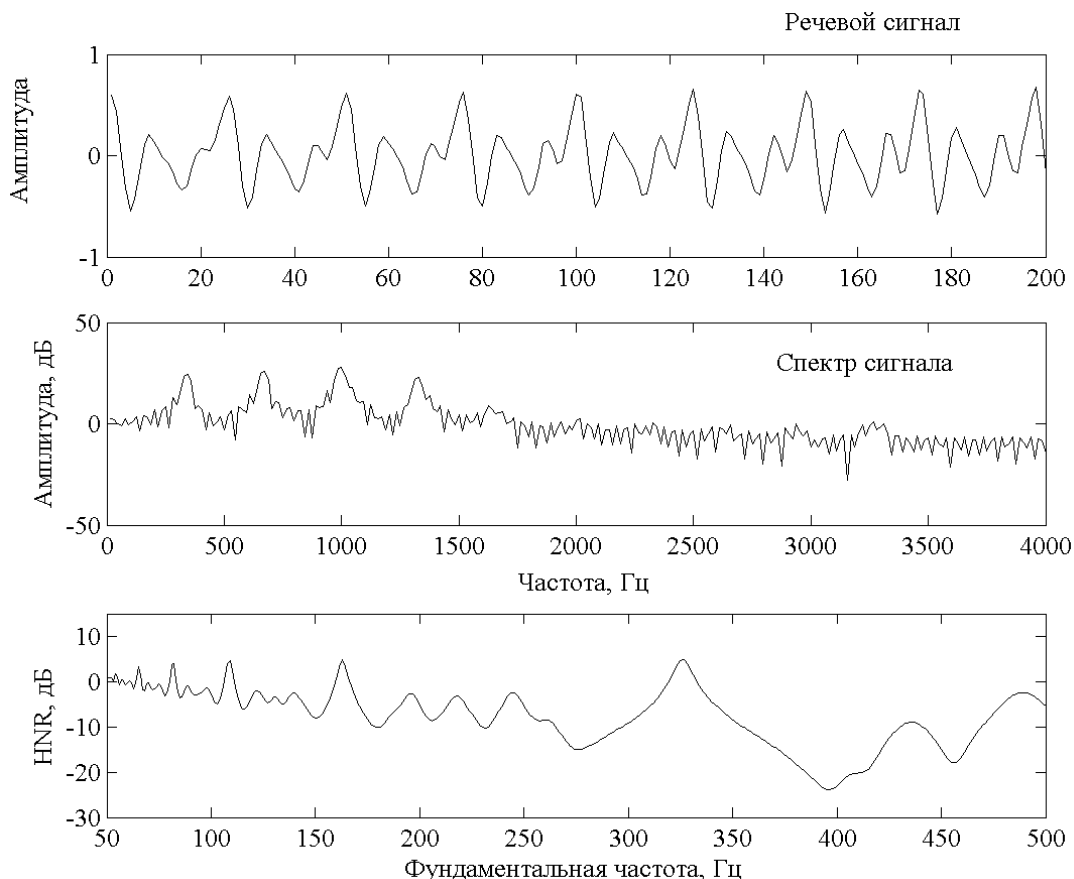


Рис. 12.8 – Речевого сигнала, его спектр и зависимость отношения “гармоники / шум” от предполагаемого значения фундаментальной частоты

Шаг 2. Найти максимумы НАКФ и отобрать из соответствующих им значений периода тона кандидатов.

$$p_c = \arg[\max(\psi(p))], \quad \psi(p_c) \geq 0,3 \cdot \psi_{\max}(p), \quad (12.23)$$

где $\psi_{\max}(p)$ – максимальное значение НАКФ на данном сегменте.

Шаг 3. Для каждого кандидата рассчитать функцию стоимости с учетом прошлой информации о контуре частоты основного тона.

$$D_{i,j} = d_{i,j} + \min_{k \in I_{i-1}} \{D_{i-1,k} + \delta_{i,j,k}\}, \quad (12.24)$$

где: $d_{i,j}$ – локальная стоимость j -го кандидата в момент времени i , $\delta_{i,j,k}$ – стоимость перехода от k -го кандидата в момент времени $i-1$ к j -му кандидату в момент времени i , $1 \leq j \leq I_i$; I – количество кандидатов.

Шаг 4. Выбрать в качестве предварительной оценки периода основного тона на анализируемом сегменте кандидат j с минимальной стоимостью $D_{i,j}$.

Шаг 5. Если сегмент вокализован, провести процедуру расчёта параметров гармонической компоненты и одновременного уточнения значения частоты основного тона.

Процедура расчёта параметров гармонической компоненты и одновременного уточнения значения частоты основного тона включает в себя следующие действия:

- 1) рассчитывается коэффициент затухания по формуле (12.6);
- 2) осуществляется поиск максимума отношения “гармоники / шум” в окрестности значения F_0^{est} . На каждой итерации рассчитываются векторы амплитуд A и фаз θ гармоник по формулам (12.8) – (12.10) с последующим синтезом гармонической компоненты по формуле (12.5) и определением нового значения HNR .

Шаг 6. Если сегмент вокализован, провести процедуру расчёта параметров гармонической компоненты и одновременного уточнения значения отклонения частоты основного тона. Здесь осуществляется поиск максимума отношения “гармоники / шум” в окрестности нулевого значения данного отклонения. На каждой итерации рассчитываются векторы амплитуд A и фаз θ гармоник по формулам (12.8) – (12.10) с последующим синтезом гармонической компоненты по формуле (12.5) и определением нового значения HNR .

Условием оптимальности будет следующее выражение:

$$\Delta F_0^{opt} = \arg \max \left(HNR(F_0^{opt} + \Delta F_0) \right), \quad \Delta F_{0l} \leq \Delta F_0 \leq \Delta F_{0r}, \quad (12.25)$$

где диапазон $[\Delta F_{0l}; \Delta F_{0r}]$ представляет возможную область значений изменения частоты основного тона.

Ниже приведён пример реализации данной методики на псевдоязыке:

```
If  $NRG > ThrNRG$  then
begin
  For  $k = P_{min}$  to  $P_{max}$  do  $NACF(k)$ ;
   $Candidates = Search\_max\_NACF(ThrAd)$ ;
   $F_0' = Track\_DP(Candidates)$ ;
  if  $F_0' > 0$  then
  begin
     $Beta = Compute\_Beta(F_0')$ ;
    for  $j = 1$  to  $n\_iterations$  do
       $[F_0^{opt}, A, \theta] = Golden\_Section(F_{0l}(j), F_{0r}(j))$ ;
  end;
end;
```

В приведённом алгоритме используются следующие обозначения: NRG – энергия сигнала на сегменте; $ThrNRG$ – заданное пороговое значение энергии вокализованного сегмента, определённое эмпирически; $NACF(k)$ – подпрограмма расчета НАКФ по формуле (12.22); $Search_max_NACF(ThrAd)$ – подпрограмма поиска максимумов нормализованной автокорреляционной функции (НАКФ), с которыми отождествляются кандидаты частоты основного тона $Candidates$. Рассматриваются только значения максимумов НАКФ, превышающие некоторый адаптивно изменяющийся порог $ThrAd$; $Track_DP(Candidates)$ – подпрограмма выбора

траектории частоты основного тона методом динамического программирования [22]. Помимо кандидатов частоты основного тона в ней рассматривается и гипотеза о некокализированности речевого сегмента. В качестве параметров в данной подпрограмме используются значения НАКФ, с которыми отождествлены кандидаты частоты основного тона *Candidates*, а также расстояния между кандидатами частоты основного тона для смежных сегментов. Траектория частоты основного тона должна представлять собой для вокализованных звуков плавную линию. Результатом работы подпрограммы является приблизительная оценка фундаментальной частоты $F'_0 > 0$ или вывод о некокализированности данного сегмента речи $F'_0 = 0$; *Compute_Beta*(F'_0) – подпрограмма, осуществляющая расчёт фактора изменения гармонических амплитуд β ; *Golden_Section*($F_{0l}(j), F_{0r}(j)$) – подпрограмма, осуществляющая поиск максимума отношения “гармоники / шум” в окрестности значения F'_0 методом “золотого сечения”, выполняется $n_iterations$ раз. На каждой итерации в подпрограмме осуществляется расчёт векторов гармонических амплитуд A и фаз θ по формулам (12.8) – (12.10) с последующим синтезом гармонической компоненты по формуле (12.5) и определением нового значения HNR .

Итак, результатом работы представленного алгоритма является набор параметров, характеризующий гармоническую компоненту речевого сигнала: фундаментальная частота F_0 , изменение фундаментальной частоты ΔF_0 , вектор гармонических амплитуд A и вектор гармонических фаз θ , фактор β .

На рис. 12.9 показан результат работы приведенного алгоритма в составе схемы обработки речевого сигнала, представленной на рис. 12.2.

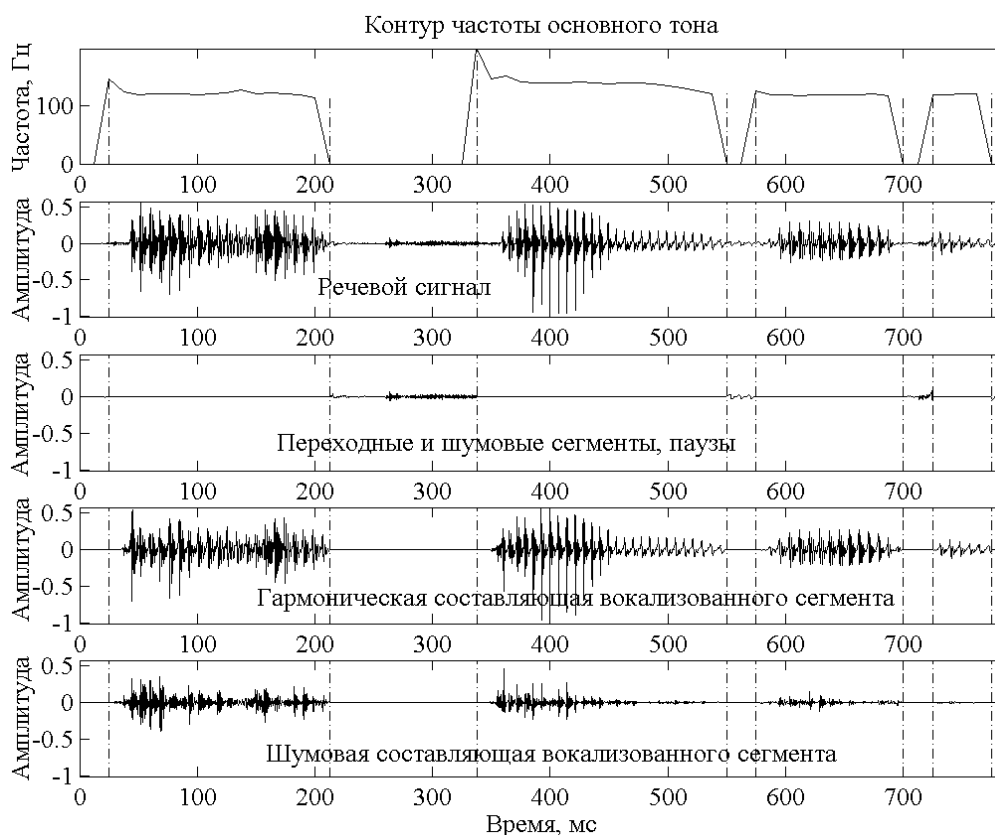


Рис. 12.9 – Сепарация речевого сигнала

12.3.5 Анализ переходных сегментов

Как уже было отмечено, отрезки речи, носящие переходный характер или содержащие какие-либо локальные явления, резко влияющие на стационарность, не могут быть описаны в рамках модели “гармоники+шум” с достаточной степенью точности. С другой стороны, возможны ошибки классификации сегмента алгоритмом определения частоты основного тона, что также повлияет на качество реконструкции речи. Для повышения качества анализа такие случаи выделяются в отдельный класс, вводится модель переходных сегментов.

Обработка данных типов речевых отрезков обычно осуществляется во временной области. Задача по определению параметров может решаться с помощью таких подходов как метод “согласованной подгонки” [23], возможен также какой-либо метод кодирования формы сигнала, например CELP [12].

12.4 Синтез речи по параметрам модели “гармоники + шум”

Процесс синтеза речи по параметрам модели сводится к суммированию сгенерированных гармонических компонент периодической составляющей с найденными в процессе анализа амплитудами, фазами и частотами с реконструированной аperiodической составляющей.

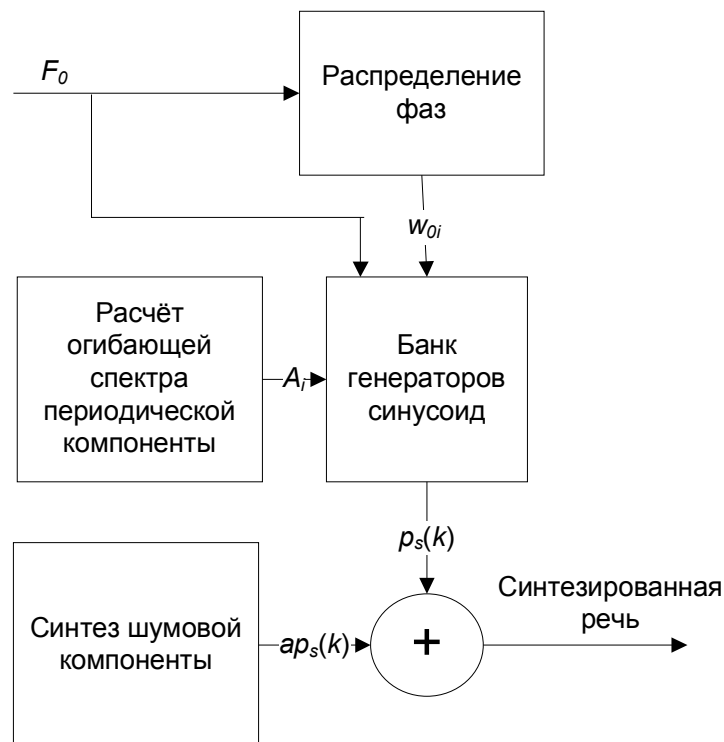


Рис. 12.10 – Схема синтеза речи по модели “гармоники + шум”

Синтез речевого сигнала с использованием модели “гармоники + шум” осуществляется по схеме, показанной на рис. 12.10 и включает следующие шаги:

1) на основе траектории фундаментальной частоты выполняется распределение фаз гармоник основного тона. При этом, во-первых, необходимо обеспечить отсутствие реверберации в синтезируемой речи и, во-вторых, минимизировать разрывы фаз между соседними ок-

нами. Начальные фазы периодических компонент первого фрейма речевого сообщения задаются случайным образом, однако начальные фазы последующих фреймов задаются в соответствии с формулой:

$$\varphi_k(m) = \text{mod} \left(\left(2\pi m F_{0,k} N \Delta t + \frac{2\pi m (F_{0,k} - F_{0,k-1}) N \Delta t}{2} + \varphi_{k-1}(m) \right), 2\pi \right), \quad (12.26)$$

где $\varphi_k(m)$ – начальная фаза m -ой гармоники во фрейме сигнала k , $F_{0,k}$ – частота основного тона речевого сигнала в k -ом фрейме.

2) По известным амплитудам для предыдущего “ $r - 1$ ”-го и текущего “ r ”-го фреймов интерполированная $\tilde{A}_l^r(n)$ может быть найдена с помощью линейной интерполяции:

$$\tilde{A}_l^r(n) = A_l^{r-1} + (A_l^r - A_l^{r-1}) \left(\frac{n}{N_p} \right), l = \overline{1, L_s}, n = \overline{1, N_p}, \quad (12.27)$$

где \tilde{A}_l^r – амплитуда l -ой гармоники из “ r ”-го фрейма; A_l^{r-1} – амплитуда l -ой гармоники из “ $r - 1$ ”-го фрейма; N_p – длина фрейма синтеза.

3) банком синусоидальных генераторов, частота которых линейно меняется на протяжении окна, формируется тональная составляющая.

4) синтезируется шумовая компонента. Ее спектр находится взвешиванием спектра белого шума огибающей спектра шумовой компоненты.

5) складываются тональная и шумовая компоненты.

12.5 Гибридный вокодер со скоростью передачи 2,4/4,8 кбит/с, основанный на декомпозиции речевого сигнала на периодическую и аperiodическую компоненты

В основе любой системы кодирования речи лежит идея о достижении высокого качества реконструированной речи при минимальной скорости битового потока. Одной из современных тенденций при разработке средств сжатия является закрытие какого-то диапазона скоростей потока одним семейством кодеков [24]. Такой подход предоставляет пользователю гибкий выбор между качеством речи и полосой пропускания.

Общая схема системы сжатия отображена на рис. 12.11.

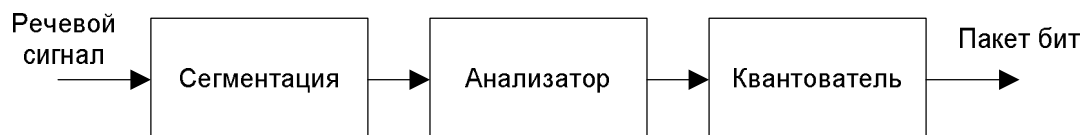


Рис. 12.11 – Общая схема кодера

Устройство сегментации выделяет речевой отрезок для последующего анализа. Размер речевого отрезка обычно определяется алгоритмом работы анализатора. В описываемом кодере речевой сигнал анализируется на сегменте длительностью 25 мс с перекрытием 50%. Такая длительность окна анализа позволяет осуществить достаточно точное определение параметров при небольшом объеме бит.

Работа анализатора основана на отдельной обработке периодической и аperiodической компонент речевого сигнала. Для декомпозиции речи на эти составляющие применяется дискретное преобразование Фурье, согласованное с изменением частоты основного тона.

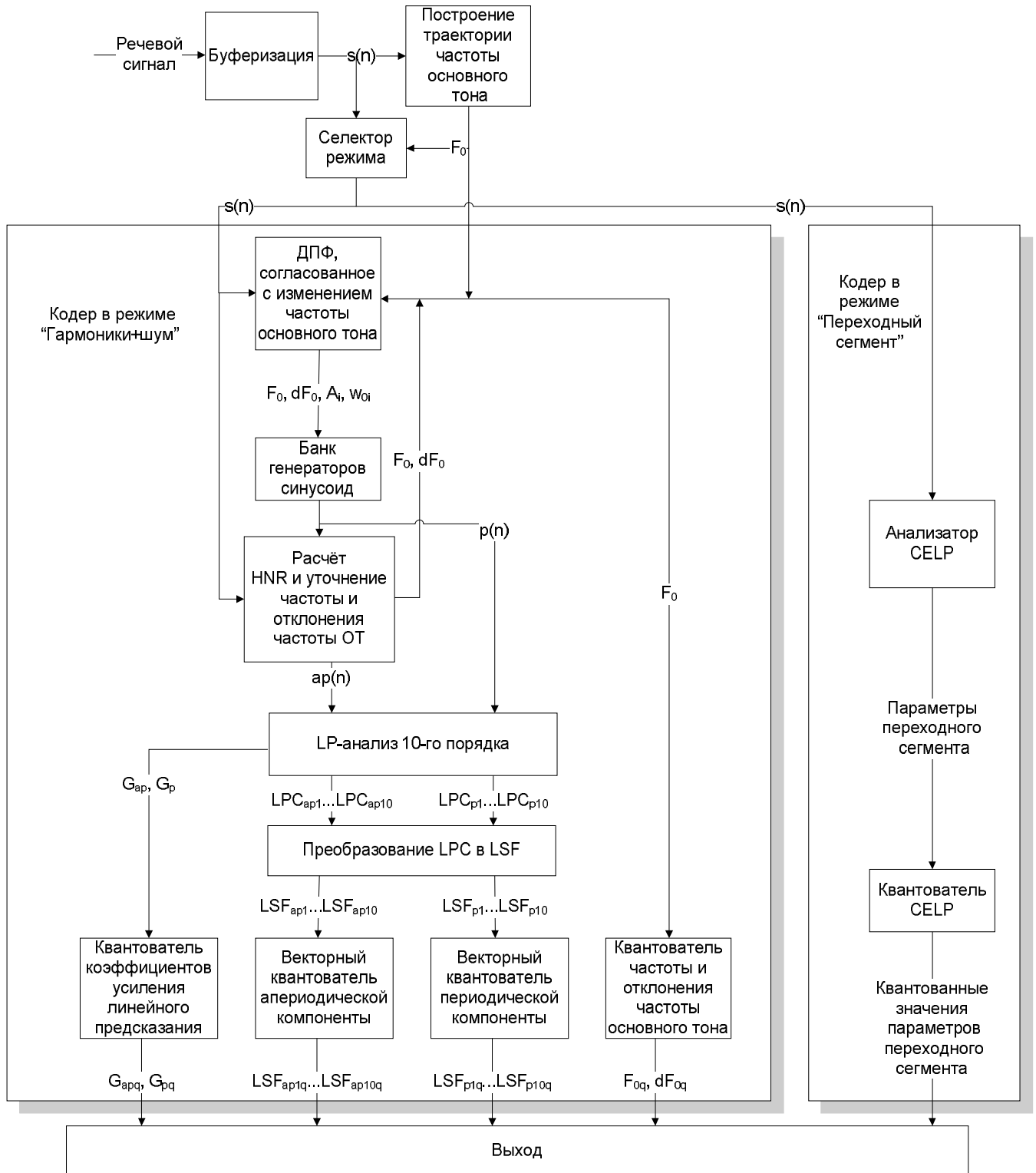


Рис. 12.12 – Структурная схема кодера 4,8 кбит/с

Для анализа переходных сегментов была принята CELP-модель анализа-синтеза речевого сигнала [25] для скорости 4,8 кбит/с, поскольку при данном подходе в синтезированную речь вносится шум, который при скоростях до 4 кбит/с маскируется на переходных сегмен-

тах [1], и LPC-модель вокального тракта [26] для скорости 2,4 кбит/с, в этом случае определяющим фактором является скорость передачи.

Подробные структурные схемы систем сжатия, основанных на декомпозиции речевого сигнала, отображены на рис. 12.12 и

рис. 12.13.

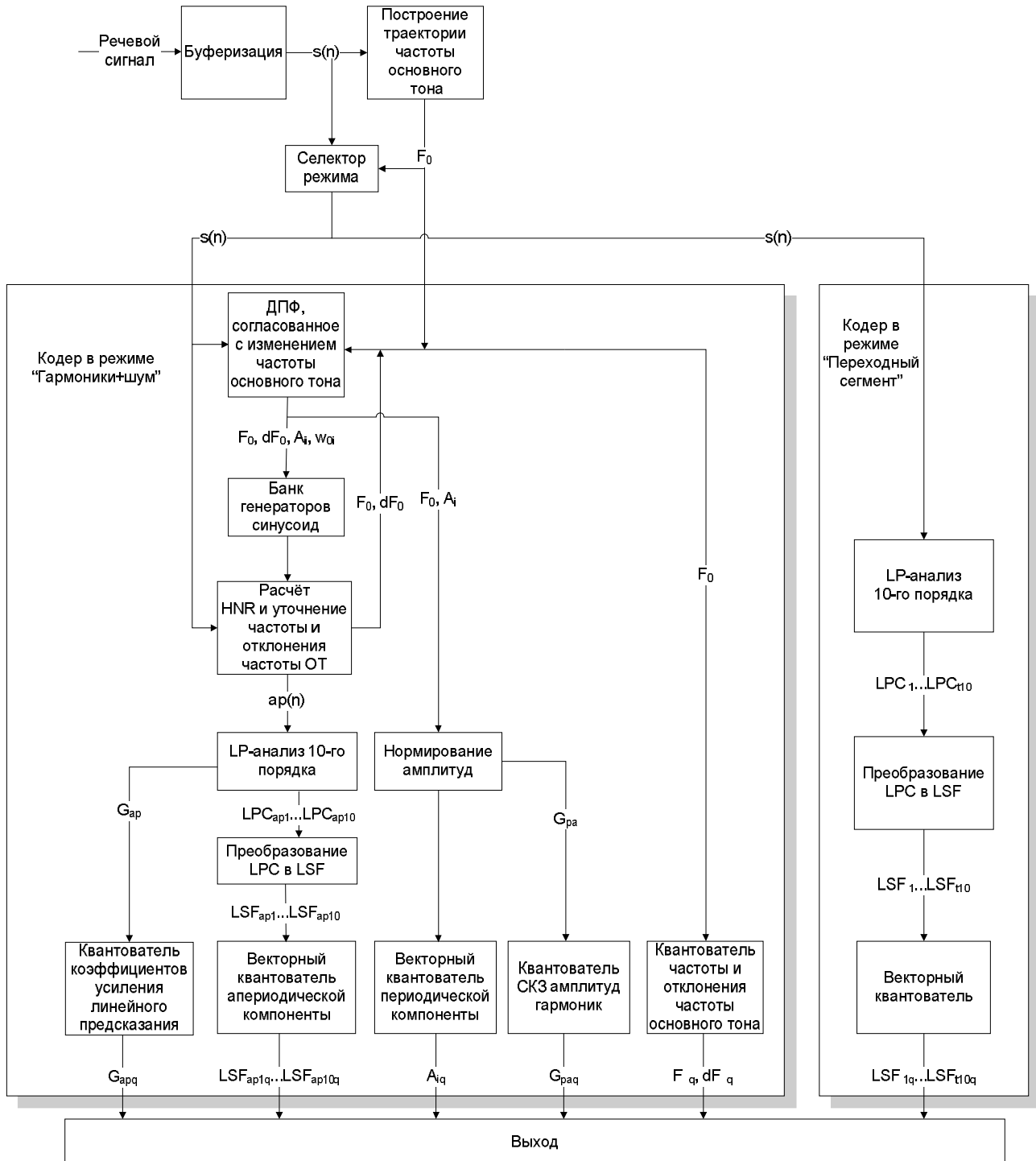


Рис. 12.13 – Структурная схема кодера 2,4 кбит/с

Периодическая компонента характеризуется частотой основного тона и спектральной огибающей, шумовая – только спектральной огибающей. Учитывая то, что шумовая компонента с точки зрения восприятия человеком гораздо менее важна, чем гармоническая, параметры

шумовой компоненты передаются один раз на сегмент. Спектральная огибающая шумовой компоненты представляется вектором коэффициентов линейного предсказания LPC_{ap} и, после преобразования в линейные спектральные пары LSF_{ap} , подвергается квантованию с расщеплением вектора SVQ (Split Vector Quantization) [27].

Тренировка кодовых книг осуществлялась на речевых отрезках, содержащих выделенную в соответствии с вышеописанным методом шумовую компоненту речевого сигнала.

В вокодере 4,8 кбит/с спектральные огибающие периодической компоненты представляются также коэффициентами линейного предсказания LPC_p и квантуются аналогичным способом. В вокодере 2,4 кбит/с квантуются непосредственно векторы амплитуды гармоник периодической компоненты [28], которые предварительно нормируются на величину среднеквадратичного значения (СКЗ). Тренировка кодовых книг осуществлялась на речевых отрезках, содержащих выделенную в соответствии с вышеописанным методом гармоническую компоненту речевого сигнала. Параметры спектральной огибающей в вокодере 4,8 кбит/с передаются для каждого субфрейма (два раза на сегмент), в вокодере 2,4 кбит/с – только для одного субфрейма, параметры второго субфрейма восстанавливаются в декодере с помощью линейной интерполяции.

Табл. 12.1 – Распределение бит для системы сжатия речевого сигнала со скоростью потока 2,4 кбит/с в режиме “гармоники+шум”.

Параметр модели	Кол-во парам./фрейм	Кол-во бит/фрейм	Кол-во фреймов/с	Скорость потока, бит/с
$\{A_i\}$	2 вектора из переменного числа элементов	0+11	40	440
G_{ap}	2	0+5		200
F_0	2	0+8		320
dF_0	2	0+4		160
LSF_{ap}	1 вектор из 10 элементов	23		920
G_{ap}	1	7		280
$Model$	1	1		40
$Reserve$	1	1		40
Результирующий поток данных:				2400

Табл. 12.2 – Распределение бит для системы сжатия речевого сигнала со скоростью потока 4,8 кбит/с в режиме “гармоники+шум”.

Параметр модели	Кол-во парам./фрейм	Кол-во бит/фрейм	Кол-во фреймов/с	Скорость потока, бит/с
LSF_p	2 вектора из 10 элементов	27+27	40	2160
G_p	2	5+5		400
F_0	2	8+8		640
dF_0	2	4+4		320
LSF_{ap}	1 вектор из 10 элементов	23		920
G_{ap}	1	7		280
$Model$	1	1		40
$Reserve$	1	1		40
Результирующий поток данных:				4800

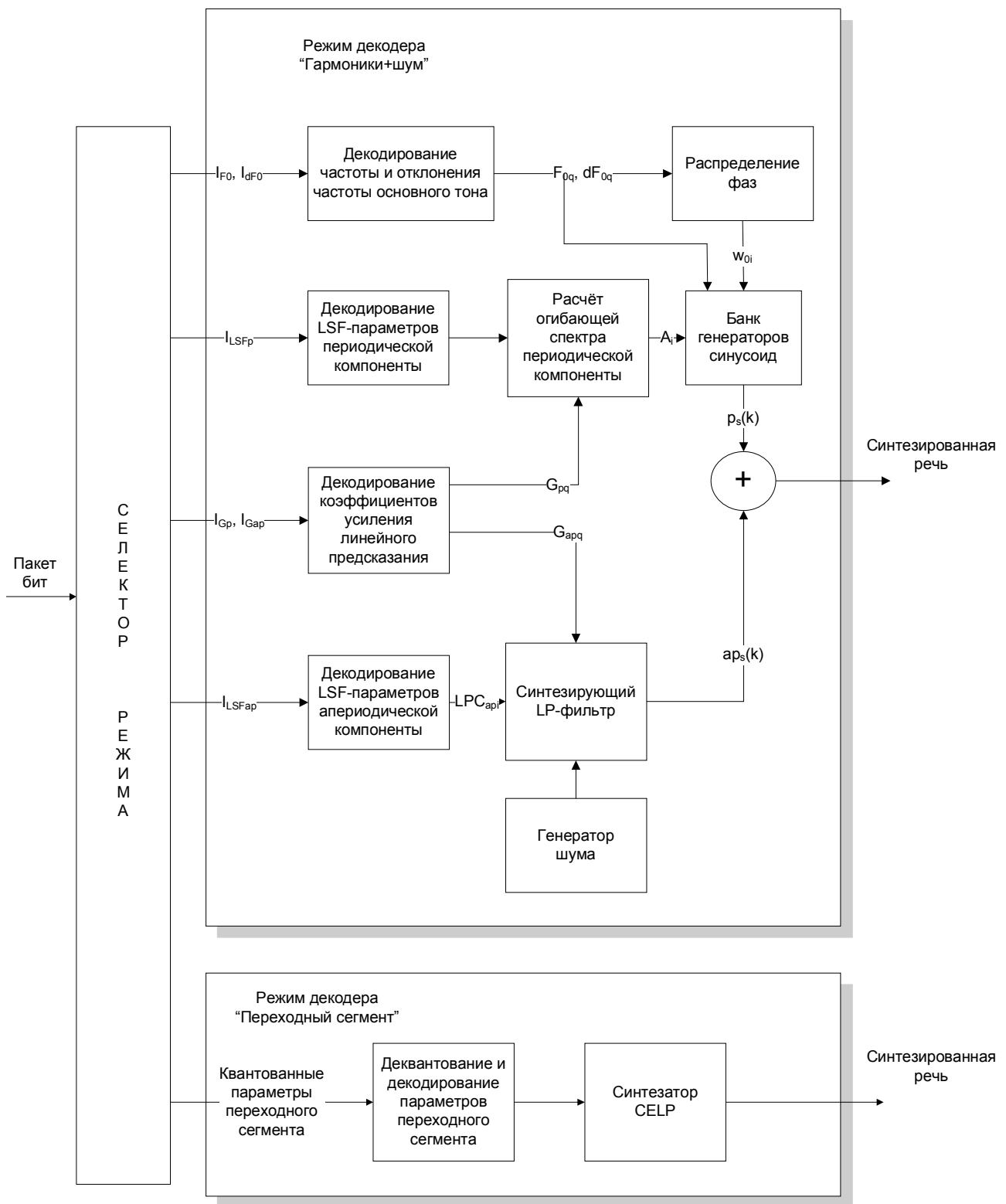


Рис. 12.14 – Структурная схема декодера 4,8 кбит/с

Таким образом, в канал связи из системы сжатия речевого сигнала для случая вокализованного сегмента передаются следующие параметры:

– для скорости 2,4 кбит/с: индекс кодовой книги амплитуд гармоник, индекс кодовой книги линейных спектральных пар аperiodической компоненты, индекс кодовой таблицы среднеквадратического значения амплитуд гармоник, индекс кодовой таблицы коэффициентов усиления линейного предсказания аperiodической компоненты, индекс кодовой таблицы частоты основного тона и отклонения частоты основного тона;

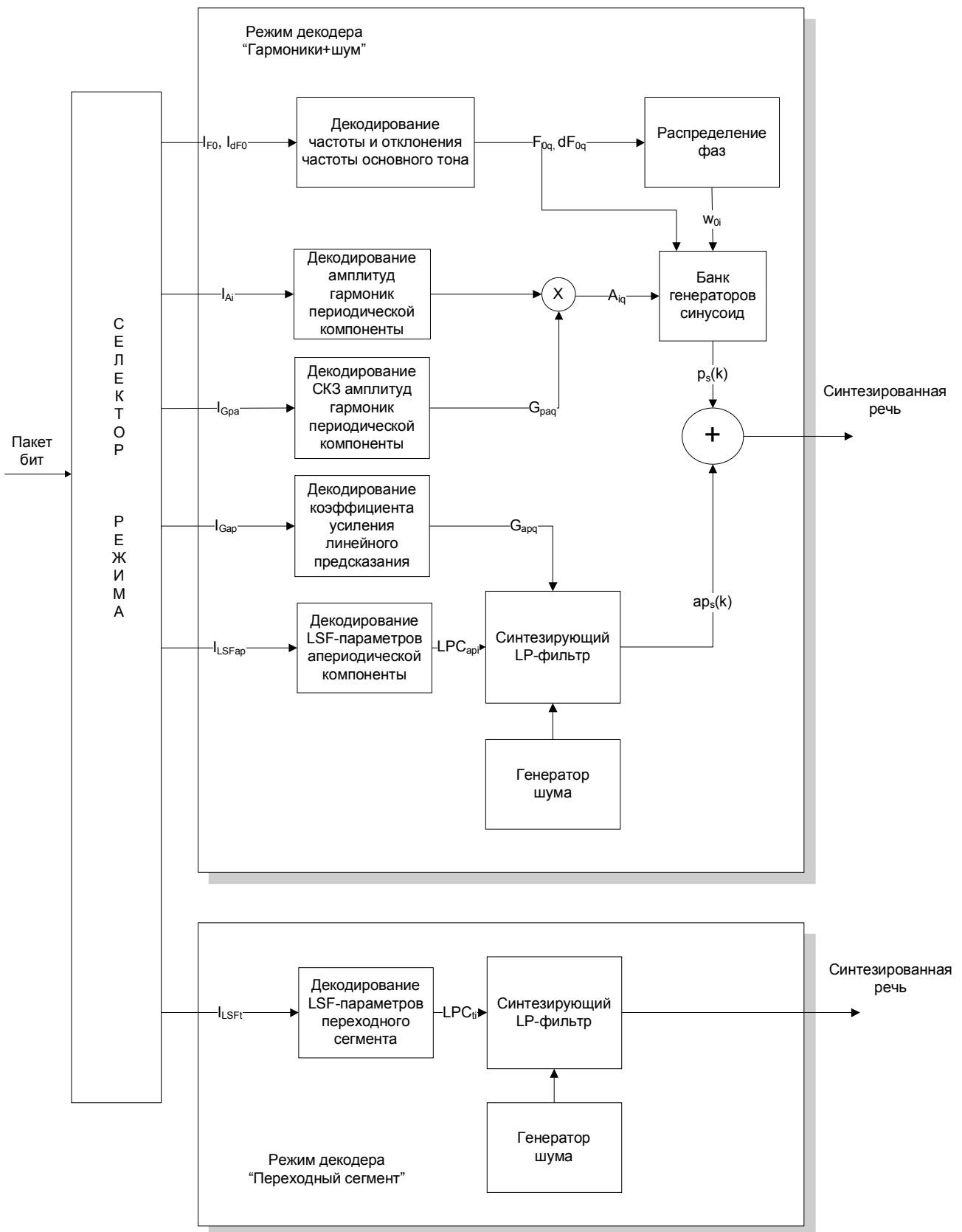


Рис. 12.15 – Структурная схема декодера 2,4 кбит/с

– для скорости 4,8 кбит/с: индексы кодовых книг линейных спектральных пар периодической компоненты для двух субфреймов и аperiodической компоненты для всего сегмента, индексы кодовых таблиц коэффициентов усиления линейного предсказания периодической компоненты для двух субфреймов и аperiodической компоненты для всего сегмента, индексы

кодовой таблицы частоты основного тона и отклонения частоты основного тона для двух суб-фреймов.

Вне зависимости от скорости передачи, в канал связи передаётся также признак модели, согласно которой будет осуществляться синтез. Для сопряжения с внешней аппаратурой 1 бит был зарезервирован. Распределение бит для системы сжатия речевого сигнала со скоростью потока 2,4 кбит/с приведено в табл. 12.1 а для системы сжатия со скоростью потока 4,8 кбит/с – в табл. 12.2.

На рис. 12.14 и рис. 12.15 показаны подробные структурные схемы систем декомпрессии речи. На вход поступают индексы соответствующих кодовых таблиц. После деквантования речевых параметров и восстановления характеристик спектральных огибающих, осуществляется синтез речи в соответствии с вышеописанным принципом.

Метод гибридного кодирования речи интегрирует преимущества гармонических кодеров и кодеров формы сигнала. Неформальные тесты прослушивания показали, что гибридная модель позволяет избавиться от ограничений, налагаемых в методах, которые построены на базе какой-то одной модели. При определении параметров гармонической модели использовался математический аппарат дискретного преобразования Фурье, согласованного с изменением контура частоты основного тона, то обеспечивает высокую точность результатов при условии корректного определения фундаментальной частоты. Применение техники динамического программирования в совокупности с последующим одновременным уточнением частоты основного тона и определением параметров гармонической модели в цикле с обратной связью методом “анализ-через-синтез” позволяет добиться качественного разделения речевого сигнала на периодическую (непосредственно гармоническую) и аperiodическую (шумовую) компоненты. Сепарация речи на эти две разные по своей природе составляющие, а, следовательно, описание их разными наборами параметров, приводит к уменьшению пространства состояний для процедуры квантования, что, в свою очередь, упрощает подготовку кодовых книг и улучшает их качество.

12.6 Список использованных источников

1. Kondoz A.M., Digital speech: coding for low bit rate communication systems. New York: John Wiley & Sons Ltd., 2004, 441 p.
2. Das A., Gersho A., Low-rate multimode multiband spectral coding of speech. Int. J. of speech technology, 1999, Vol. 2, pp. 317–327.
3. Almeida L.B., Tribolet J.M., Harmonic coding: a low bit-rate, good quality, speech coding technique. Acoustics, speech and signal processing: proceedings of IEEE int. conf. (ICASSP-82), Paris, France, May 3–5, 1982, pp. 1664–1667.
4. McAulay R.J., Quatieri T.F., Speech analysis/synthesis based on a sinusoidal representation. IEEE Trans. on Acoust., Speech and Signal Processing, 1986, Vol. ASSP-34, pp. 744–754.
5. Griffin D.W., Lim J.S., Multiband excitation vocoder, IEEE Transactions on Acoustics, Speech and Signal processing, 1988, Vol. ASSP-36, pp. 1223–1235.

-
6. Spanias A.S., Speech coding: a tutorial review, Proc. IEEE, 1994, Vol. 82, No. 10, pp. 1541–1582.
 7. Serra X., A system for sound analysis / transformation / synthesis based on a deterministic plus stochastic decomposition: PhD thesis, Stanford, 1989, 178 p.
 8. Bailly G., Accurate estimation of sinusoidal parameters in an harmonic + noise model for speech synthesis. Speech communication and technology: proceedings of 6-th europ. conf. (EUROSPEECH'99), Budapest, Hungary, September 5–9, 1999, pp. 1051-1054.
 9. Etemoglu C.O. and et al., Speech Coding with an Analysis-by-Synthesis Sinusoidal Model, Acoustics, speech and signal processing: proceedings of IEEE int. conf. (ICASSP'2000), Istanbul, Turkey, June 5–9, 2000, pp. 1371–1374.
 10. Серков В.В., Петровский А.А., Использование закономерностей психоакустики при низкоскоростном кодировании речи. Цифровая обработка сигналов и ее применение: труды 3-й междунар. науч.-техн. конф. (DSPA'2000), Москва, Россия, 29 ноября – 1 декабря 2000, с.241–244.
 11. Stylianou Y., Applying the harmonic plus noise model in concatenative speech synthesis, IEEE Trans. on Speech and Audio Proc., 2001, Vol. 9, No. 1, pp. 21–29.
 12. Shlomot E., Cuperman V. and Gersho A., Hybrid coding: combined harmonic and waveform coding of speech at 4 kb/s, IEEE Trans. on Speech and Audio Proc., 2001, Vol. 9, No.6, pp.632–646.
 13. d'Alessandro C., Darsinos V, Yegnanarayana B. Effectiveness of a periodic and aperiodic decomposition method for analysis of voice sources. IEEE Trans. on Speech and Audio Proc., 1998, Vol.6, No. 1, pp. 12–23.
 14. Yegnanarayana B., d'Alessandro C., Darsinos V., An iterative algorithm for decomposition of speech signals into periodic and aperiodic components, IEEE Trans. on Speech and Audio Proc., 1998, Vol.6, No. 1, pp. 1–11.
 15. Jackson P., Shadle C., Pitch-scaled estimation of simultaneous voiced and turbulence-noise components in speech, IEEE Trans. on Speech and Audio Proc., 2001, Vol.9, No. 7, pp. 713–726.
 16. Petrovsky A., Zubricki P, Sawicki A., Tonal and noise separation based on a pitch synchronous DFT analyzer as a speech coding method, The proc. of the 16th European conference on Circuits theory and design, ECCTD'03, Sep. 1-4, 2003, Vol.III, Cracow, Poland, pp.169–172.
 17. Sercov V., Petrovsky A., An improved speech model with allowance for time-varying pitch harmonic amplitudes and frequencies in low bit-rate MBE coders, Speech communication and technology: proceedings of 6-th europ. conf. (EUROSPEECH'99), Budapest, Hungary, September 5–9, 1999, pp. 1479–1482.
 18. Петровский А., Серков В., Низкоскоростной вокодер с моделью речеобразования «гармоники + шум», Цифровая обработка сигналов, 2002, №2, с.2–12.
 19. Jensen, J., Jensen, S.H., Hansen, E., Harmonic exponential modeling of transitional speech segments, Acoustics, speech, and signal processing: proceedings of IEEE int. conf. (ICASSP'2000). Istanbul, Turkey, June 5–9, 2000, Vol. 3, pp. 1439–1442.
-

-
20. Zubricki P., Pavlovec A., Petrovsky A, Analysis-by-synthesis parameters estimation in the harmonic coding framework by pitch tracking modified DFT, *New trends in audio and video*, Dobrucki A., Petrovsky A. and Skarbek W. Eds. Byalystok 2006, pp.233–246.
 21. George E.B., Smith M.J.T., Analysis-by-synthesis/overlap-add sinusoidal modeling applied to the analysis and synthesis of musical tones, *Journal of the Audio Engineering Society*, 1992, Vol. 40, No. 6, pp. 497–515.
 22. Talkin D., Robust algorithm for pitch tracking, *Speech Coding and Synthesis*, Kleijn, W.B. and Palival, K.K. Eds. Elsevier, Amsterdam, Netherlands, 1995.
 23. Mallat, S. and Zhang, Z., Matching pursuit with time-frequency dictionaries, *IEEE Trans. on Sig. Proc.*, 1993, Vol. 41, No. 12, pp. 3397–3415.
 24. Wang T., Koishida K., Cuperman V., Gersho A. and Collura J., A 1200/2400 bps Coding Suite Based on MELP, 2002 IEEE Workshop on Speech Coding, Tsukuba, Japan, 6-9 October 2002, pp.90–92.
 25. Schroeder M., Atal B., Code-excited linear prediction (CELP): high-quality speech at very low bit rates, *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 10, 1985, pp.937–940.
 26. Tremain T., The Government Standard Linear Predictive Coding Algorithm: LPC-10, *Speech Technology Magazine*, April 1982, pp.40–49.
 27. Gersho A., Gray R., *Vector quantization and signal compression*, Boston, Kluwer Academic Publishers, 1992.
 28. Павловец А., Петровский А., Использование закономерностей психоакустики в процедуре квантования параметров гармонической модели речевого сигнала, *Речевые технологии*, 2008, №4, с.50–60.

13 ТЕКСТОЗАВИСИМАЯ КОНВЕРСИЯ ГОЛОСА НА ОСНОВЕ ДЕКОМПОЗИЦИИ РЕЧЕВОГО СИГНАЛА НА ГАРМОНИКИ, ШУМ И ПЕРЕХОДНЫЕ ФРЕЙМЫ

Павловец А.Н., Лившиц М.З., Лихачёв Д.С., Петровский А.А.

АННОТАЦИЯ

Представлена система конверсии голоса, основанная на модели сепарации речевого сигнала на «гармоники+шум» и переходные фреймы с раздельной конверсией каждой компоненты модели. Преимущество системы конверсии голоса в данном случае складывается из достоинств анализа – синтеза гармонической модели и анализа – конверсии переходных сегментов во временной области. Неформальные тесты прослушивания показали, что узнаваемость диктора соответствует приблизительно 70%, реконструированная речь характеризуется достаточно высокой разборчивостью.

Ключевые слова – гармоники плюс шум, конверсия голоса, переходные сегменты, сепарация речевого сигнала.

СОДЕРЖАНИЕ

13.1	Задача конверсии голоса.....	285
13.2	Декомпозиция речевого сигнала на гармоническую составляющую, шум и переходные фреймы	287
13.2.1	Селектор режима	287
13.2.2	Гармонический анализ.....	287
13.2.3	Анализ переходных сегментов	288
13.3	Реализация системы конверсии голоса	288
13.3.1	Особенности процесса обучения.....	288
13.3.2	Процесс динамической временной трансформации	289
13.3.3	Процесс конверсии голоса.....	295
13.4	Экспериментальные результаты.....	296
13.5	Список использованных источников.....	298

13.1 Задача конверсии голоса

Проблема конверсии голоса становится очень популярной в мире. Актуальность данной темы обусловлена широким применением устройств конверсии голоса в мультимедиа-системах реального времени: синтез речи по тексту (устранение «компьютерного акцента»); виртуальное дублирование (восстановление звуковых дорожек кинофильмов); защита свидетелей (применение в судебной практике); оперативная смена диктора в коммуникационных системах (озвучивание SMS-сообщений в мобильных телефонах) [1].

Одной из критических задач в области конверсии речи является задача определения параметров конверсии. Для решения этой задачи предлагается метод, основанный на деком-

позиции речевого сигнала на 3 составляющие (гармоническую, шумовую и переходную) с идеей раздельной их обработки и дальнейшей конверсии.

Конверсия голоса представляет собой процесс изменения характеристик речи одного диктора (исходного) таким образом, чтобы она звучала так, как будто была произнесена другим диктором (целевым) [1]. Общий алгоритм процесса конверсии показан на рис. 13.1.

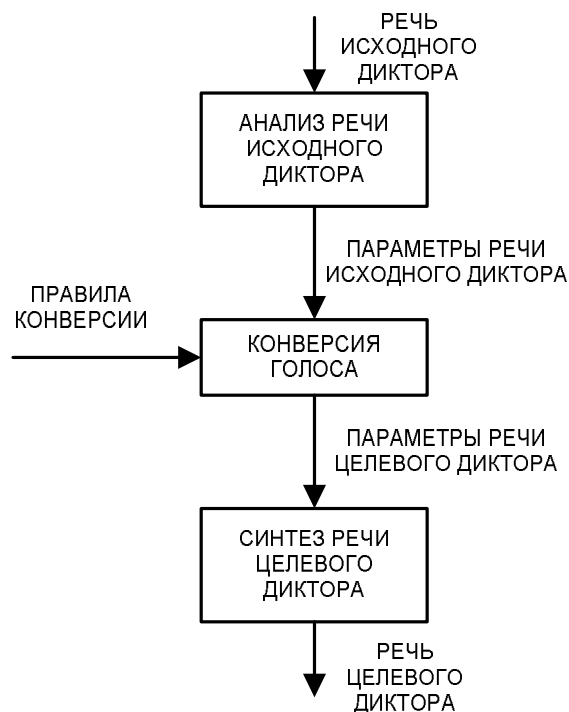


Рис. 13.1 – Типовая схема процесса конверсии голоса

Процесс конверсии голоса можно разбить на два этапа. На первом из них (этапе обучения) выделяется множество характеристических параметров исходного и целевого дикторов, и определяются правила конверсии, посредством которых параметры исходного диктора будут преобразовываться в параметры целевого диктора. На втором этапе (этапе конверсии) характеристики речи исходного диктора преобразуются с использованием правил, определенных на первом этапе.

При реализации системы конверсии голоса требуется решить две основные проблемы: как и какие параметры извлекать из речевого сигнала, подлежащего преобразованию, и как модифицировать эти параметры таким образом, чтобы преобразованная речь была похожа на речь целевого диктора.

Важной задачей в системах конверсии голоса является определение параметров голоса. В работе [1] был улучшен подход [2,3], который был основан на модели “гармоники плюс шум”, путём использования декомпозиции анализируемой речи на вокализованную и шумоподобную компоненты. Ранее подобный метод был применён в области кодирования речи [4,5].

Дальнейшие исследования показали, что модель [1] не является достаточной в полной мере, поскольку с её помощью нельзя корректно анализировать переходные сегменты речи.

Это послужило причиной дополнения модели [1] режимом анализа переходных сегментов [6,7].

Для модификации параметров речи было предложено большое количество статистических подходов. Популярность приобрели методы, основанные на векторном квантовании. В этом случае определение правил конверсии представляет собой установление соответствия между кодовыми книгами, представляющими акустические классы дикторов. В [8] был описан метод, основанный на жесткой кластеризации и дискретном соответствии между кодовыми книгами. Получаемый характеристический вектор y_t в момент времени t определяется путём квантования исходного характеристического вектора и подстановки вместо него соответствующей центроиды из кодовой книги целевого диктора.

Жесткая кластеризация подразумевает значительную ошибку квантования. В данном случае используется модификация метода из [8]. В зависимости от типа сегмента речи используются различные кодовые книги. Целью является обеспечение лучшего качества конверсии голоса посредством двухрежимной системы.

13.2 Декомпозиция речевого сигнала на гармоническую составляющую, шум и переходные фреймы

13.2.1 Селектор режима

Многорежимный подход был успешно опробован в области кодирования речи и аудио, например [9,10]. Первая из упомянутых работ представляет гибридный кодер речи, который сочетает параметрический кодер, работающий в частотной области (для случаев стационарной вокализованной и стационарной невокализованной речи), с кодером формы сигнала, работающим во временной области (для переходных сегментов). Вторая работа использует сегментацию аудиосигнала на три различных сигнала: сигнал, моделирующий синусоидальную составляющую, сигнал, который моделирует все переходные сегменты, и шумовой сигнал.

В предлагаемой модели речь рассматривается как:

$$s(i) = h(i) + r(i), \quad (13.1)$$

где $h(i)$ – вокализованная (гармоническая) компонента, $r(i)$ – сигнал-остаток (шумоподобная компонента или переходный сегмент).

Режим работы речевого анализатора полностью определяется наличием либо отсутствием основного тона (рис. 13.2).

13.2.2 Гармонический анализ

Так же как в работах [11] и [12], в подходе, представляемом здесь, считается, что вокализованная и шумоподобная составляющие присутствуют во всём диапазоне частот. Спектральный анализ проводится в области гармоник фундаментальной частоты речи, для этого ДПФ было модифицировано таким образом, чтобы учитывать изменение её контура. Точность определения параметров модели, а именно частоты основного тона, гармонических амплитуд и фаз, повышена с использованием метода анализа через синтез. Предполагается, что гармо-

ническая компонента определяется суммой гармоник фундаментальной частоты с изменяющимися во времени амплитудами и фазами. Декомпозиция выполняется во временной области, шумоподобная компонента определяется разностью между оригинальной речью и синтезированной гармонической компонентой [6].

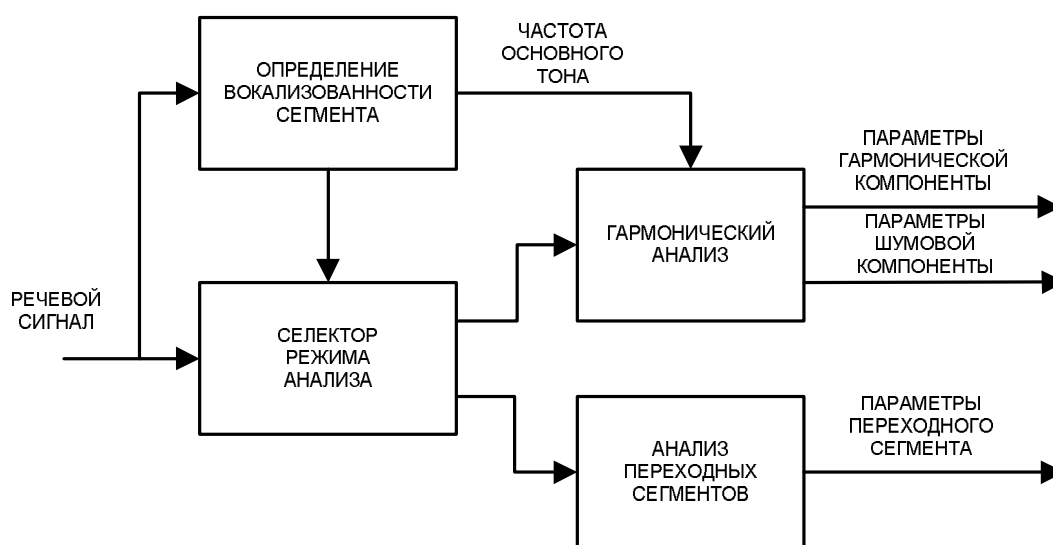


Рис. 13.2 – Схема декомпозиции речи

13.2.3 Анализ переходных сегментов

Для анализа переходных сегментов был использован метод ACELP в соответствии с рекомендациями ITU-T G.729 [13].

13.3 Реализация системы конверсии голоса

13.3.1 Особенности процесса обучения

В работах [1,14] была сделана попытка решения проблемы конверсии голоса с помощью таких методов преобразования спектра как сопоставление кодовых книг гармонических амплитуд [14] и преобразование линейных спектральных частот (Line Spectral Frequencies – LSF) с помощью модели гауссовых смесей [1]. Принимая во внимание возможность использования двухрежимного анализатора речи, естественным было бы использовать различные функции конверсии для огибающих спектра, получаемых в каждом режиме. Таким образом, для каждого режима работы был применен улучшенный метод отображения кодовых книг [15], но тренировка системы конверсии голоса осуществлялась отдельно для каждого режима.

В дальнейшем будет рассмотрен процесс тренировки функций конверсии и алгоритм работы системы конверсии голоса, основанной на модели речи “гармоники+шум+переходы” (ГШП).

Для тренировки системы конверсии в каждом режиме необходимо подготовить два множества параметров целевого и исходного дикторов. После проведения гармонического анализа для процесса конверсии используются такие параметры как спектральная огибающая, представленная LSF-коэффициентами, и фундаментальная частота F_0 . Анализ переходных сег-

ментов, осуществляемый с помощью метода ACELP [13], предоставляет для модификации LSF-коэффициенты фильтра, моделирующего вокальный тракт, период основного тона T_0 , коэффициенты усиления последовательностей адаптивного и фиксированного возбуждения G_a и G_f соответственно.

В ходе фазы обучения для каждого режима анализатора рассчитываются два множества параметров для исходного и целевого дикторов, в дальнейшем для каждого режима определяется свой набор параметров конверсии.

Для осуществления преобразования таких параметров как фундаментальная частота F_0 , период основного тона T_0 , коэффициенты усиления последовательностей адаптивного и фиксированного возбуждения G_a и G_f используется метод линейного преобразования математического ожидания и дисперсии, при этом предполагается, что математические ожидания этих параметров содержат существенную часть информации, специфической для каждого диктора. Предполагается также, что значения параметров каждого диктора подчиняются распределению Гаусса и имеют характерные средние значения и отклонения.

Обозначив модифицируемый параметр как P , можно определить линейное преобразование следующим образом:

$$P^t = \frac{\delta_t}{\delta_s} (P^s - \mu_s) + \mu_t, \quad (13.2)$$

где P^t , P^s – один из параметров целевого и исходного дикторов соответственно, δ_s , μ_s , δ_t , μ_t – среднеквадратическое отклонение и математическое ожидание соответствующего параметра исходного и целевого дикторов соответственно.

Как уже упоминалось, данные для конверсии спектра представляют собой два множества пар спектральных векторов (LSF), полученных с помощью подхода, который будет описан далее. В ходе обучения для установления более точного сопоставления спектральных векторов исходного и целевого дикторов использовался алгоритм динамической временной трансформации (DTW – Dynamic Time Warping) [16]. На рис. 13.3 показана фаза тренировки кодовой книги спектральных векторов.

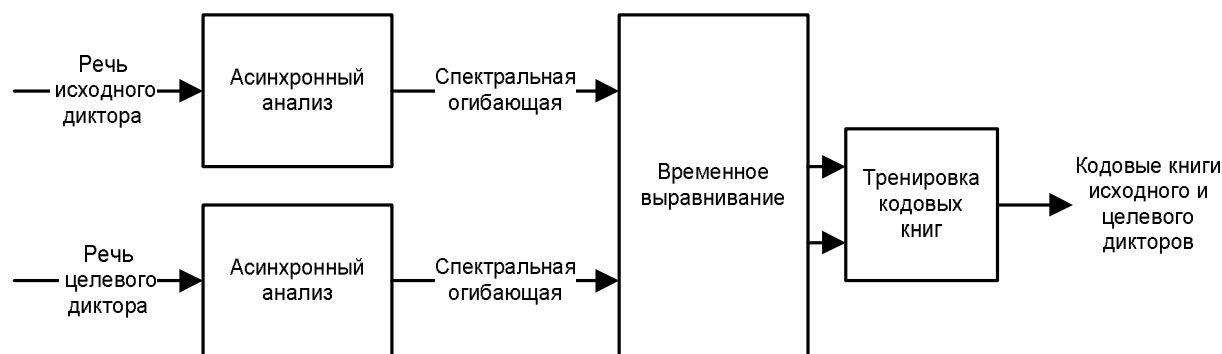


Рис. 13.3 – Схема тренировки кодовых книг

13.3.2 Процесс динамической временной трансформации

Предположим, имеется две последовательности наборов LSF-параметров: для целевого диктора – LSF^{tag} и для исходного – LSF^{source} . Необходимо выполнить выравнивание дан-

ных двух последовательностей по времени, т.е. таким образом сопоставить их по длине путём вставки или удаления элементов из LSF^{source} , чтобы общее среднеквадратическое отклонение между элементами этих последовательностей стремилось к минимальному значению.

Общее отклонение (расстояние) между последовательностями LSF^{tag} и LSF^{source} можно определить как:

$$D(LSF^{tag}, LSF^{source}) = \sum_{s=1}^{N_p} d(p_s), \quad (13.3)$$

где $d(p_s)$ – расстояние между LSF_n^{source} и LSF_m^{tag} , N_p – количество элементов в пути.

Тогда процесс нахождения оптимального пути сводится к минимизации общего отклонения:

$$P = \arg \min_p [D(LSF^{tag}, LSF^{source})], \quad (13.4)$$

Таким образом, задача сводится к нахождению функции выравнивания (пути):

$$P = p_1, \dots, p_s, \dots, p_{N_p}, \quad p_s = (n_s, m_s), \quad (13.5)$$

каждое значение которой показывает, какой элемент последовательности LSF^{source} следует удалить, а какой вставить (рис. 13.4), чтобы достигнуть минимального значения общего отклонения.

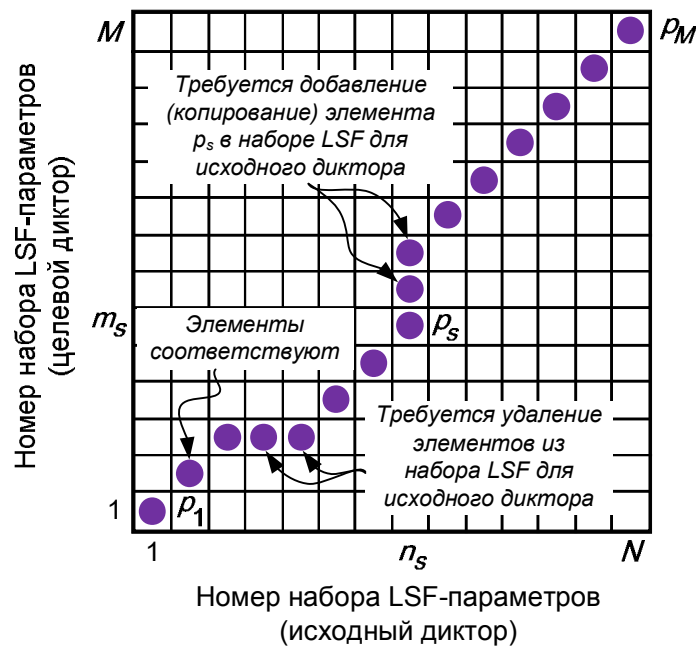


Рис. 13.4 – Иллюстрация алгоритма динамической временной трансформации

Алгоритм для нахождения функции выравнивания (оптимального пути) может быть описан следующим образом:

Шаг 1. Вычислить матрицу локальных среднеквадратических отклонений d , каждый элемент которой является евклидовым расстоянием между двумя соответствующими элементами последовательностей LSF^{source} и LSF^{tag} , и определяется по следующему выражению:

$$d(n, m) = \sqrt{\sum_k^{10} [LSF_n^{source}(k) - LSF_m^{tag}(k)]^2}, \quad n = \overline{1, N}, \quad m = \overline{1, M}, \quad (13.6)$$

Шаг 2. Вычислить матрицу весов D каждый элемент которой характеризует вклад соответствующего элемента матрицы d в общее среднеквадратическое отклонение.

Шаг 2.1. Положим начальное условие $D(1,1) = d(1,1)$. Вычислить первую строку матрицы D :

$$D(n, 1) = D(n - 1, 1) + d(n, 1), \quad n = \overline{1, N}. \quad (13.7)$$

Шаг 2.2. Вычислить первый столбец матрицы:

$$D(1, m) = D(1, m - 1) + d(1, m), \quad m = \overline{1, M}. \quad (13.8)$$

Шаг 2.3. Далее двигаясь по матрице d слева направо снизу вверх вычисляются следующие элементы матрицы D :

$$D(n, m) = \min[D(n, m - 1), D(n - 1, m - 1), D(n - 1, m)] + d(n, m), \quad n = \overline{1, N}, \quad m = \overline{1, M}. \quad (13.9)$$

В процессе вычисления для каждой ячейки матрицы запоминается индекс соседней ячейки, которая вносит минимальный вклад в общую ошибку.

Шаг 3. Анализируя матрицу D в направлении от $D(N, M)$ до $D(1, 1)$ и учитывая определённые на предыдущих этапах индексы ячеек, которые вносят меньший вклад в общее отклонение по сравнению с соседними, определяется наилучший путь $P = p_1, \dots, p_s, \dots, p_M$ с точки зрения минимизации величины общего отклонения.

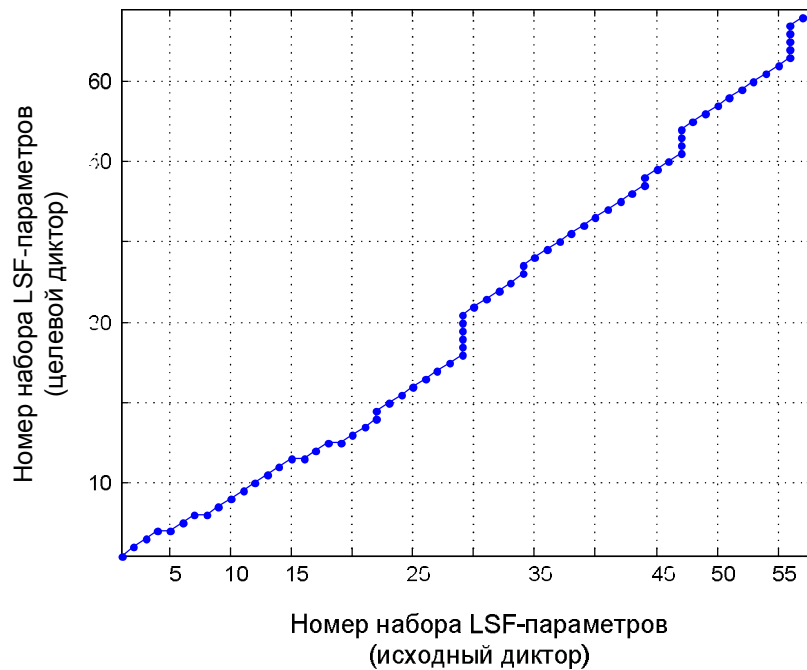


Рис. 13.5 – Функция сопоставления для двух наборов LSF-параметров исходного и целевого мужских дикторов с длиной $N = 57$ и $M = 68$ соответственно

Полученный в результате работы алгоритма путь $P = p_1, \dots, p_s, \dots, p_M$ является функцией сопоставления для обрабатываемых последовательностей, которая показывает какие элементы необходимо удалить из исходной последовательности, а какие добавить.

Например, на рис. 13.5 отображена функция сопоставления для двух наборов LSF-параметров исходного и целевого мужских дикторов с длиной $N = 57$ и $M = 68$ соответственно. Данным наборам соответствует фраза “Испорченный контакт”.

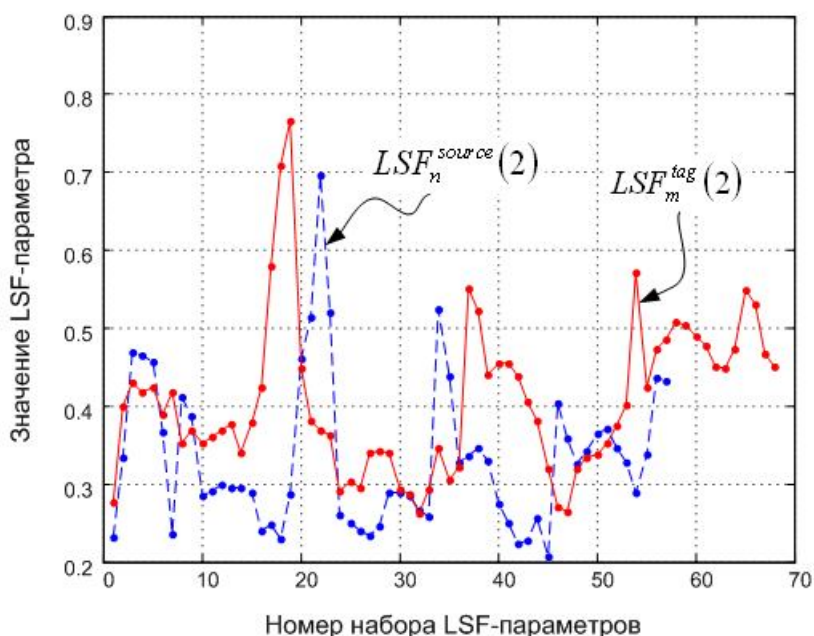


Рис. 13.6 – Последовательности $LSF_n^{source}(2)$ и $LSF_m^{tag}(2)$ до сопоставления

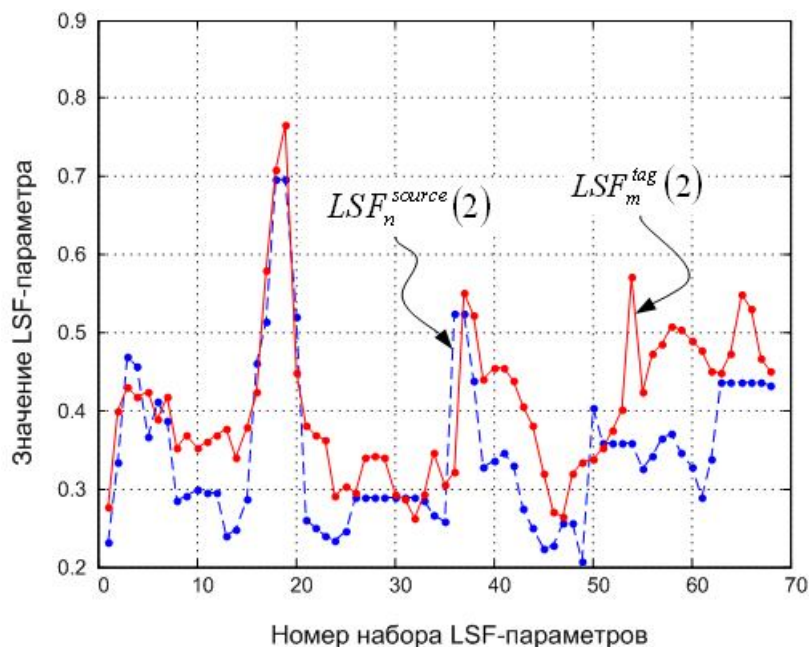


Рис. 13.7 – Последовательности $LSF_n^{source}(2)$ и $LSF_m^{tag}(2)$ после сопоставления

На рис. 13.6 отображены последовательности $LSF_n^{source}(2)$ и $LSF_m^{tag}(2)$ до сопоставления, а на рис. 13.7 эти же последовательности – после сопоставления. На рис. 13.8 отображены

последовательности $LSF_n^{source}(5)$ и $LSF_m^{tag}(5)$ до сопоставления, а на рис. 13.9 – сопоставленные.

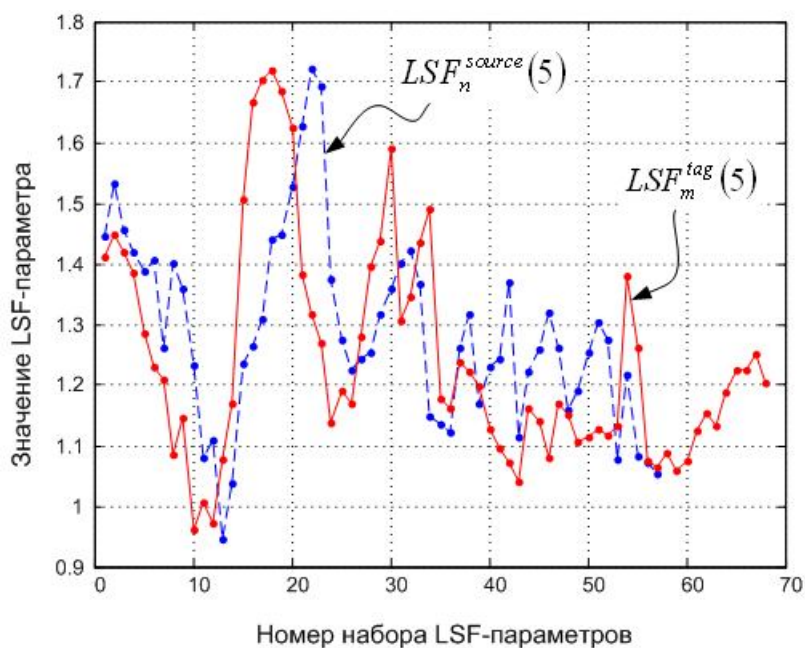


Рис. 13.8 – Последовательности $LSF_n^{source}(5)$ и $LSF_m^{tag}(5)$ до сопоставления

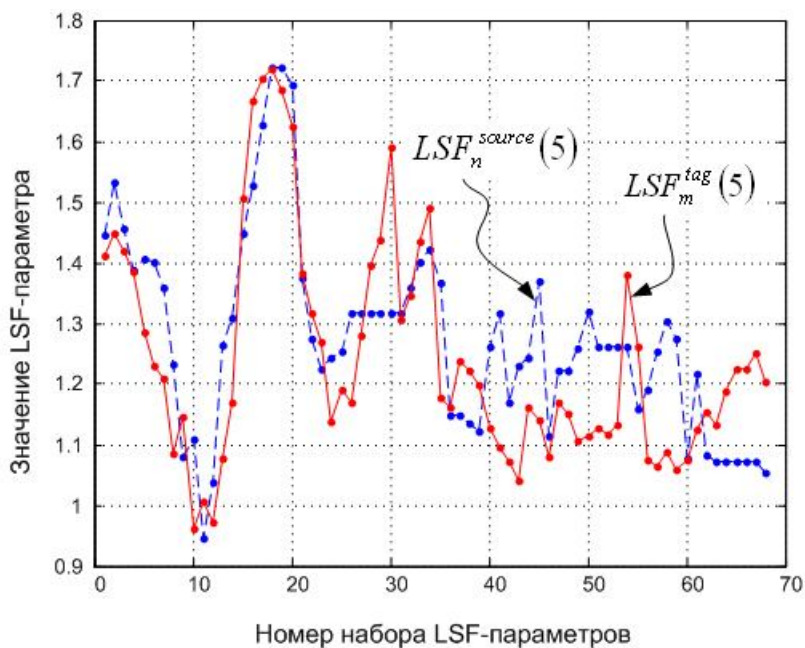


Рис. 13.9 – Последовательности $LSF_n^{source}(5)$ и $LSF_m^{tag}(5)$ после сопоставления

Ниже приводится программный код в Matlab:

```
% Загрузка последовательности LSF для исходного диктора
load LSFx lsfx;
% Загрузка последовательности LSF для целевого диктора
load LSFy lsfy;
N=length(lsfx);
M=length(lsfy);
```

```

% Шаг 1. Вычисление матрицы локальных
& среднеквадратических отклонений d(i,j)
for i=1:N
    for j=1:M
        d(i,j)=sqrt(sum((lsfx(i,:)-lsfy(j,:)).^2));
    end;
end;
% Инициализация матрицы весов D
D = zeros(N,M);
D(1:N, 1:M) = d;

% Шаг 2.1. Вычисление первой строки матрицы D
D(1,1)=d(1,1);
D(2:N,1)=D(1:N-1,1)+d(2:N,1);

% Шаг 2.2. Вычисление первого столбца матрицы D
D(1,1:M)=D(1,1:M)+d(1,1:M);

% Шаг 2.2. Вычисление первого столбца матрицы D
iRow(1,1:M) = NaN;
iCol(1:N+1,1) = NaN;
iCol(1) = 0;
D=[iRow;D];
D=[iCol D];
N=N+1;
M=M+1;
% Вспомогательная матрица
pt = zeros(N,M);

% Шаг 2.3. Вычисление следующих элементов матрицы D
for i = 2:N;
    for j = 2:M;
        [dmin, ind] = min([D(i-1, j-1), D(i-1, j), D(i, j-1)]);
%        [dmin, tb] = min([D(i, j), D(i, j+1)]);
        D(i,j) = D(i,j)+dmin;
        pt(i-1,j-1) = ind;
    end;
end;

% Шаг 3. Анализ матрицы D и определение оптимального пути
i = N-1; j = M-1; p = i; q = j;
while (i >=0) && (j > 1)
    ind = pt(i,j);
    if (ind == 1)
        i = i-1; j = j-1;
    elseif (ind == 2)
        i = i-1;
    elseif (ind == 3)
        j = j-1;
    else
        error;
    end
    p = [i,p];
end

```

```

q = [j,q];
end;
% Формирование новой последовательности x_new
x_new=[];
for i=1:length(p)-1
    if q(i)~=q(i+1)
        x_new=[x_new;lsfx(p(i),:)] ;
    end;
end;
x_new=[x_new;lsfx(p(length(p)),:)] ;
% Предполагается, что последовательность LSF целевого диктора
% остаётся неизменной
y_new=lsfy;
% Отображение оптимального пути на графике
figure();
plot(p,q,'.b','MarkerSize',20); hold on;
plot(p,q,'-b');
axis([1 N 1 M]);
xlabel('p');
ylabel('q'); grid on; zoom on;

```

13.3.3 Процесс конверсии голоса

После определения параметров конверсии, сам процесс конверсии выполняется как показано на рис. 13.10. На вход системы поступает речевой сигнал, оцифрованный с частотой дискретизации 8 кГц.

Прежде всего, детектор голосовой активности, реализованный в соответствии с [17] проверяет сегмент входного сигнала на наличие речи. В случае если сегмент содержит тишину и/или фоновый шум, он передается напрямую на выход.

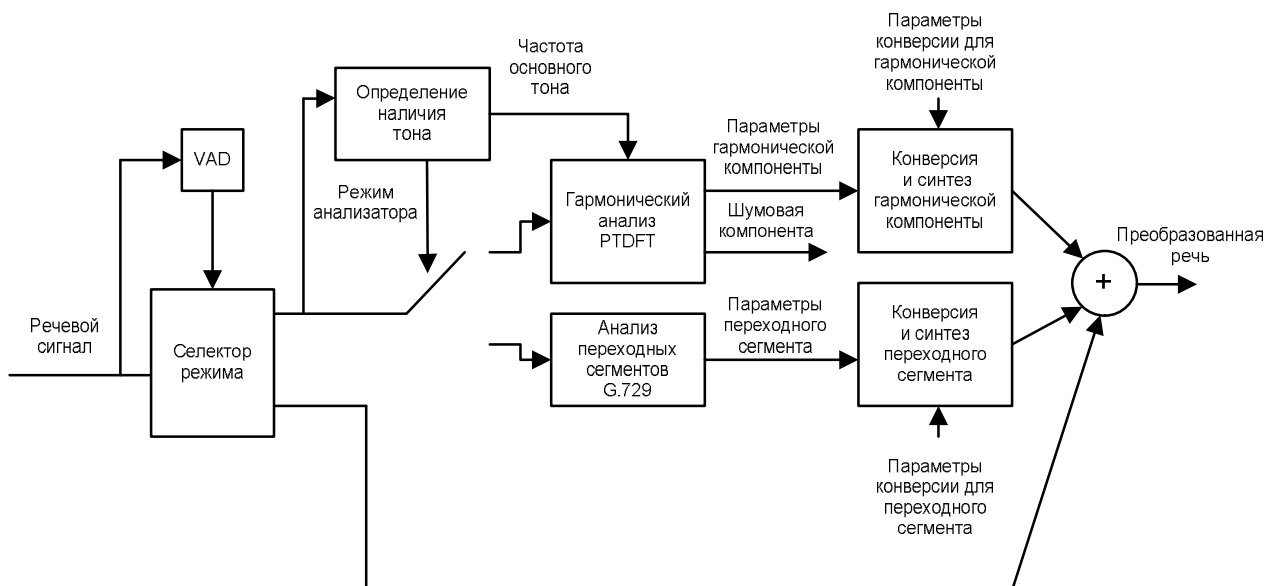


Рис. 13.10 – Схема процесса конверсии голоса

Детектор тона определяет, будет ли речевой сегмент передан для обработки в модуль гармонического анализа либо в модуль анализа переходных сегментов. Затем параметры, вы-

деленные с помощью одного из этих модулей анализа, подвергаются преобразованию, и по преобразованным параметрам осуществляется синтез сегмента речи целевого диктора.

Функция конверсии векторов LSF в обоих режимах имеет следующий вид:

$$F(x_t) = \sum_{i=1}^N p_i c_i, \quad (13.10)$$

где p_i – вес, характеризующий вероятность принадлежность вектора x_t к i -тому акустическому классу, представленному в кодовой книге размерности N центроидой c_i :

$$p_i = \frac{e^{-d_i}}{\sum_{j=1}^N e^{-d_j}}, \quad (13.11)$$

где d_i – мера искажения:

$$d_i = \sum_{k=1}^m v_k |c_i - x_t|. \quad (13.12)$$

В выражении (13.12) величина m представляет собой размерность вектора, а v_k – вес, рассчитанный по формуле обратного гармонического среднего, с помощью которого учитывается перцептуальный фактор близости смежных LSF:

$$v_k = \frac{1}{\omega_k - \omega_{k-1}} + \frac{1}{\omega_{k+1} - \omega_k}. \quad (13.13)$$

где ω_k – k -й коэффициент LSF, $\omega_0 = 0$, $\omega_{m+1} = \pi$.

13.4 Экспериментальные результаты

Качество работы предлагаемой системы конверсии голоса оценивалось с помощью неформальных тестов прослушивания. Для сравнения использовалась система конверсии, полностью основанная на модели ACELP. Тесты показали, что голос, производимый предлагаемой системой, достаточно естественный и его разборчивость выше, чем в системе, использующей ACELP. Рис. 13.12 и рис. 13.13 содержат спектрограммы примеров конверсии мужского голоса в женский. Для примера была выбрана фраза на польском языке: “Lubić szardaszowy pląs” (рис. 13.11). 15 фраз из [18] были использованы для обучения. Очевидно, что результат конверсии предлагаемой системой лучше соответствует гармонической структуре речи целевого диктора и содержит меньше шума. Неформальные тесты прослушивания показали превосходство предлагаемой системы над системой, использующей ACELP.

Преимущество системы конверсии голоса, основанной на модели речевого сигнала “гармоники+шум+переходы”, заключается в простом способе соединения преимуществ анализа-синтеза гармонической модели с анализом и конверсией переходных сегментов во временной области. Благодаря данному подходу, значительно снижается доля исходного диктора в конвертированной речи.

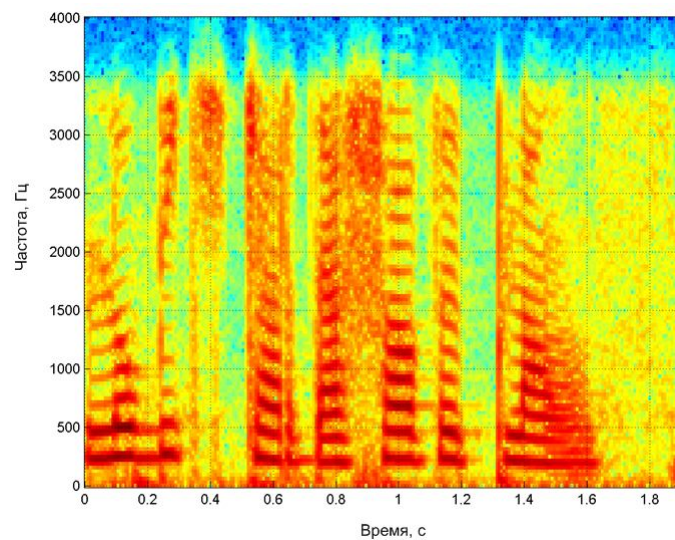


Рис. 13.11 – Спектрограмма фразы целевого диктора

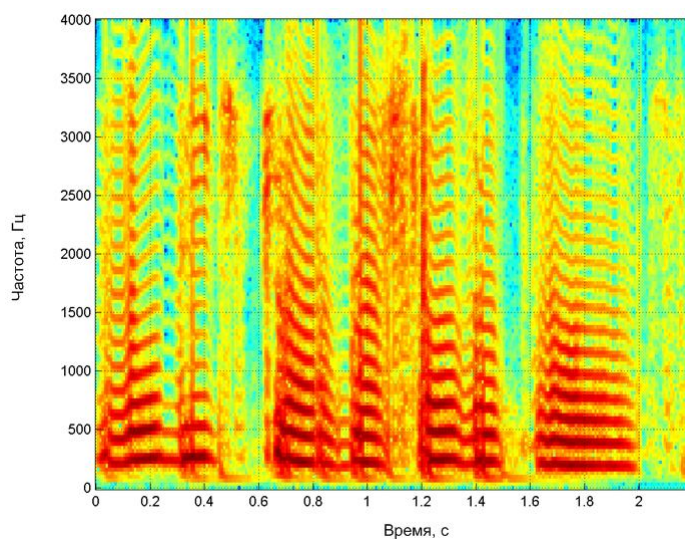


Рис. 13.12 – Пример конверсии голоса системой, основанной на модели “гармоники+шум+переходы”

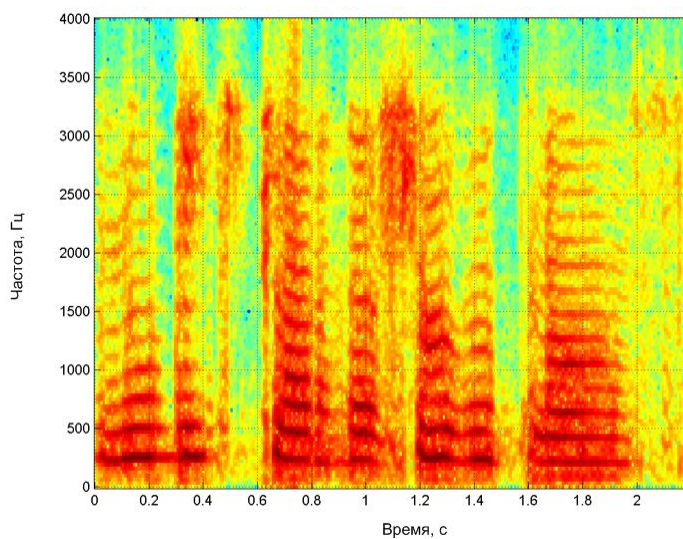


Рис. 13.13 – Пример конверсии голоса системой, основанной на ACELP-подходе

13.5 Список использованных источников

1. Moulines E., Sagisaka Y., Eds., Voice conversion: state of the art and perspectives, Speech Communication, Vol. 16, Feb. 1995.
2. Pavlovets A., Kien T., Zubricki, P., Petrovsky A., Speech analysis – synthesis based on the PTDFT for voice conversion, in Proc. of the 2007 Int. Workshop on Spectral Methods and Multirate Sig. Proc., SMMSP, Moscow, Russia, Sep. 2007, pp.203–210.
3. Stylianou Y., Laroche J., Moulines E., High-quality speech modification based on a harmonic + noise model, in Proc. of the European Conf. on Speech Communication and Technology EUROSPEECH, Madrid, Spain, Sep. 1995, pp.451–454.
4. Petrowsky A., Zubricki P., Sawicki A., Tonal and noise components separation based on a pitch synchronous DFT analyzer as a speech coding method, in Proc. European Conf. Circuit Theory and Design, Cracow, Poland, Sep. 2003, Vol. 3, pp.169–172.
5. Zubricki P., Pavlovec A., Petrovsky A., Analysis-by-synthesis parameters estimation in the harmonic coding framework by pitch tracking modified DFT, in “New trends in audio and video”, Dobrucki, A., Petrovsky, A. and Skarbek, W. Eds. Bialystok 2006, pp.233–246.
6. Павловец А., Лившиц М., Лихачёв Д., Петровский А., Конверсия голоса с использованием модели сепарации речевого сигнала на компоненты «гармоники+шум» и переходные фреймы, Речевые технологии, 2008, № 4, с.37–49.
7. Pavlovets A., Livshitz M., Lichachov D., Petrovsky A., Voice conversion based on the HNT model of speech and separate VQ learning, in Proc. of the 13-th Int. Conf. “Speech and Computer”, SPECOM, St. Petersburg, Russia, June 2009, pp.283–288.
8. Abe M., Nakamura S., Shikano K., Kuwabara H., Voice conversion through vector quantization, in Proc. of the Int. Conf. on Acoust., Speech and Sig. Proc. ICASSP, New York, USA, Apr. 1988, Vol. 1, pp.655–658.
9. Shlomot E., Cuperman V., Gersho A., Hybrid coding: combined harmonic and waveform coding of speech at 4 kb/s, IEEE Trans. on Speech and Audio Proc., Sep. 2001, Vol.9, No. 6, pp.632–646.
10. Levine S., Smith, J.O., A sines+transients+noise audio representation for data compression and time/pitch scale modifications, in Proc. 105th Conv. Audio Eng. Soc., preprint #4781, Sep. 1998.
11. Yegnanarayana B., d’Alessandro C., Darsinos V., An iterative algorithm for decomposition of speech signals into periodic and aperiodic components, IEEE Trans. on Speech and Audio Proc., Jan. 1998, Vol.6, No. 1, pp.1–11.
12. Jackson P.J.B., Shadle, C.H., Pitch-scaled estimation of simultaneous voiced and turbulence-noise components in speech, IEEE Trans. on Speech and Audio Proc., Oct. 2001, Vol. 9, No. 7, pp.713–726.
13. ITU-T Rec. G.729, Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear – prediction (CS-ACELP), Mar. 1996.

-
14. Pavlovets A., Petrovsky A., Voice conversion as a part of the voice analysis/synthesis system based on the periodic-aperiodic decomposition of speech, in Proc. of the 9th Int. Conf. on Pattern Recognition and Information Proc., PRIP, Minsk, Belarus, May 2007, Vol.2, pp.71–76.
 15. Stylianou Y., Cappe O., Moulines E., Continuous probabilistic transform for voice conversion, IEEE Trans. on Speech and Audio Processing, March 1998, Vol. 6, No. 2, pp.131–142.
 16. Huang, X., Acero A., Hon H. W., Spoken Language Processing: a guide to theory, algorithms, and system development, Prentice Hall, NJ, 2001, 980 p.
 17. ITU-T Rec. G.729, annex B, A silence compression scheme for G.729 optimized for terminals conforming to Recommendation V.70, Nov. 1996.
 18. Grocholevski S., First Database for Spoken Polish, in Proc. Int. Conf. On Language Resources and Evaluation, Grenada, 1998, pp. 1059–1062.

14 ДИСКРЕТНОЕ ПРЕОБРАЗОВАНИЕ ФУРЬЕ С НЕРАВНОМЕРНЫМ ЧАСТОТНЫМ РАЗРЕШЕНИЕМ В ПЕРЦЕПТУАЛЬНЫХ СИСТЕМАХ РЕДАКТИРОВАНИЯ ШУМА В РЕЧИ И КОМПРЕССИИ ЗВУКОВЫХ СИГНАЛОВ

Борович А., Парфенюк М., Петровский Ал.А.

АННОТАЦИЯ

Описывается новое применение ДПФ с неравномерным частотным разрешением (от англ. Warped discrete Fourier transform (WDFT)) в одноканальных перцептуальных системах редактирования шума в речи и компрессии звуковых сигналов. WDFT используется одновременно как базис для метода спектрального взвешивания так и для психоакустической модели. Обработка сигналов, осуществляемая в критических частотных полосах, более точна в контексте психоакустического моделирования по сравнению со схемами на базе ДПФ.

Ключевые слова - WDFT, психоакустика, перцептуальная обработка сигналов, компрессия, редактирование шума.

СОДЕРЖАНИЕ

14.1	Дискретное преобразование Фурье с неравномерным частотным разрешением	301
14.2	Новая схема обработки сигналов в области коэффициентов WDFT	304
14.3	Схема подавления шума в речевом сигнале на основе спектрального взвешивания	305
14.3.1	Постановка задачи	305
14.3.2	Перцептуальная система подавления шума в области коэффициентов WDFT.....	307
14.4	Оценка порога маскирования и спектральной плотности мощности шума.....	309
14.4.1	Отображение коэффициентов WDFT на критические частотные полосы	309
14.4.2	Оценка СПМ шума.....	310
14.4.3	Оценка порога маскирования.....	312
14.5	Оценка качества системы подавления шума..	316
14.6	Компрессия звуковых сигналов в области коэффициентов WDFT.....	321
14.6.1	Структура системы компрессии звуковых сигналов.....	321
14.6.2	Кодирование коэффициентов WDFT по Хаффману	326
14.6.3	Экспериментальные результаты.....	328
14.7	Список использованных источников.....	329

14.1 Дискретное преобразование Фурье с неравномерным частотным разрешением

Дискретное преобразование Фурье с неравномерным частотным разрешением (от англ. Warped discrete Fourier transform (WDFT)) [1], позволяющее получить z -преобразование конечной последовательности $x(n)$ из N входных отсчетов сигнала с неравномерным разложением коэффициентов преобразования на единичной окружности z -плоскости строго регулярно посредством использования фазового звена, определяется по следующей формуле:

$$\hat{X}(z_k) = X(\hat{z}_k) = \sum_{n=0}^{N-1} x(n)\hat{z}_k^{-n}, k = 0, \dots, N - 1, \quad (14.1)$$

где \hat{z}_k -изображения, преобразованных фазовым звеном $A(z)$, равноотстоящих точек на единичной окружности в z -плоскости, получаемые из преобразования

$$\hat{z}_k^{-1} = e^{-j\frac{2\pi k}{N}} \rightarrow \hat{z}_k^{-1} = A(z_k), k = 0, \dots, N - 1, \quad (14.2)$$

$A(z)$ устойчивое фазовое звено произвольного порядка.

Простейший вариант WDFT основан на фазовом звене первого порядка с действительным коэффициентом a

$$A(z) = \frac{z^{-1} - a}{1 - az^{-1}}. \quad (14.3)$$

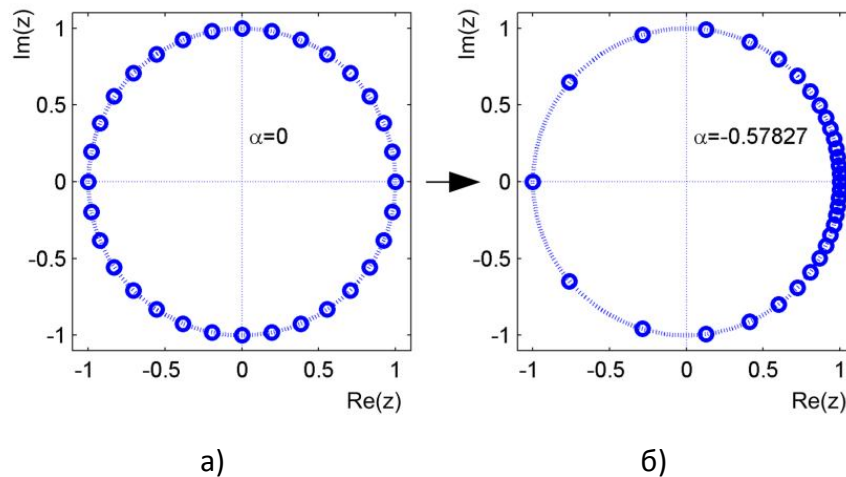


Рис. 14.1 – Расположение коэффициентов преобразования на единичной окружности в z -плоскости: ДПФ (а) и WDFT (б)

Условием стабильности является $|a| < 1$. В зависимости от знака a , растягивается низкочастотный ($a > 0$) или высокочастотный ($a < 0$) диапазон, таким образом, оставшаяся часть единичной окружности в z -плоскости становится сжатой. Формально это может быть выражено следующим образом:

$$\hat{\omega} = \omega + 2\arctan\left(\frac{a \cdot \sin(\omega)}{1 - a \cdot \cos(\omega)}\right) \text{ для } \begin{cases} z = e^{j\omega} \\ \hat{z} = e^{j\hat{\omega}} \end{cases} \quad (14.4)$$

то есть осуществляется билинейное преобразование z -плоскости в новую искривленную \hat{z} -плоскость (см. рис. 14.1).

Как обобщение ДПФ, WDFT также имеет свойства линейности, симметрии и сдвига [2,3]. Сопряженная симметрия для действительных данных имеет силу и для WDFT:

$$\hat{X}(z_{N-1-k}) = \hat{X}^*(z_k), \quad (14.5)$$

однако ряд важных свойств ДПФ потеряно.

В матричной записи (с $\hat{X}(k)$, обозначая $\hat{X}(z_k)$) WDFT может быть представлено выражением:

$$\begin{bmatrix} \hat{X}(0) \\ \hat{X}(1) \\ \vdots \\ \hat{X}(N-1) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & A(z_0) & \dots & A(z_0)^{N-1} \\ 1 & A(z_1) & \dots & A(z_1)^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & A(z_{N-1}) & \dots & A(z_{N-1})^{N-1} \end{bmatrix}}_{\mathbf{D}} \cdot \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix}. \quad (14.6)$$

В случае перцептуального деформирования оси частот матрица Вандермонда, к классу которых относится и матрица \mathbf{D} WDFT, обладают свойствами, делающими невозможным получение точной обратной матрицы. А именно, матрица WDFT может рассматриваться как сингулярная так, как между некоторыми ее строками существуют почти линейные зависимости. Существует ряд методов для аппроксимации обращений плохообусловленных матриц. В зависимости от приложения, где применяется WDFT, ошибка реконструкции речевого сигнала (вычисление обратного WDFT) может играть решающую роль на качество синтезированного сигнала (числе артефактов), например, в системах редактирования шумов окружающей среды в звуковом сигнале на основе психоакустически мотивированного правила взвешивания спектра зашумленного звука [4].

Используя матричную форму записи, вектор искажения сигнала \mathbf{d} можно определить как разность между оригинальным \mathbf{x} и реконструированным сигналами $\hat{\mathbf{x}}$:

$$\mathbf{d} = \mathbf{x} - \hat{\mathbf{x}} = (\mathbf{I} - \mathbf{D}^T \mathbf{D})\mathbf{x}. \quad (14.7)$$

Соответствующая мера в частотной области может быть выражена как средняя СПМ искажения сигнала

$$S_{dd}(\omega) = \frac{1}{N} E\{|\mathbf{e}(\omega)^H \mathbf{d}|^2\} = \frac{1}{N} \mathbf{e}(\omega)^H \mathbf{R}_{dd} \mathbf{e}(\omega), \quad (14.8)$$

где

$$\mathbf{e}(\omega) = [1 \quad e^{-j\omega 1} \quad e^{-j\omega 2} \quad \dots \quad e^{-j\omega(N-1)}]^T, \quad (14.9)$$

является вектором столбцов синусоидального базиса ДПФ, а \mathbf{R}_{dd} обозначает ковариационную матрицу сигнала ошибки.

Очевидно, что спектральное искажение (14.8), зависит от характеристик входного сигнала и качества аппроксимации обратного WDFT. Теоретически точное обращение матрицы \mathbf{D}

возможно, но данное решение будет очень не стабильным вследствие большого числа обусловленности матрицы и не найдет практического применения в системах обработки звука. Однако если удастся получить стабильное решение обратной задачи, мера величины ошибки (14.8) может быть использована для регулирования величины артефактов в синтезированном речевом сигнале.

В работе [4] было показано, что формирование избыточного не ортогонального базиса, может внести значительный вклад в коррекцию ошибки реконструкции, а именно, при подходящем выборе векторов синусоидального базиса, можно модифицировать распределение сингулярных чисел для соответствующей матрицы преобразования. Новая WDFT матрица \mathbf{D} не является квадратной и количество строк $M > N$ увеличивается. Матричное представление WDFT с избыточным базисом может быть записано следующим образом:

$$\begin{bmatrix} \hat{X}(0) \\ \hat{X}(1) \\ \vdots \\ \hat{X}(N-1) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & \hat{z}_0^{-1} & \dots & \hat{z}_0^{-N+1} \\ 1 & \hat{z}_1^{-1} & \dots & \hat{z}_1^{-N+1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \hat{z}_{M-1}^{-1} & \dots & \hat{z}_{M-1}^{-N+1} \end{bmatrix}}_{\mathbf{D}_{M \times N}} \cdot \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix}, \quad (14.10)$$

Спектральные искажения (14.8) могут быть вычислены для прямоугольной матрицы также как и для квадратной, потому что продукт $\mathbf{D}^T \mathbf{D}$ всегда $N \times N$ матрица. На рис. 14.2 показаны СПМ входного сигнала и соответствующей ошибки синтеза, вычисленные для прямоугольной матрицы $\mathbf{D}_{M \times N}$ WDFT. Как видно, ошибка синтеза уменьшается при увеличении M и может не учитываться при $M \geq M_{opt}$ (меньше -75 дБ). На практике число строк матрицы $\mathbf{D}_{M \times N}$ можно слегка уменьшить, так как перекрытие частотных характеристик полосовых фильтров выбиралось случайно и без всякого психоакустического критерия. Однако прослушивание речевых тестов показало, что спектральные искажения не слышны в реконструированном сигнале речи для $M \approx M_{opt}$.

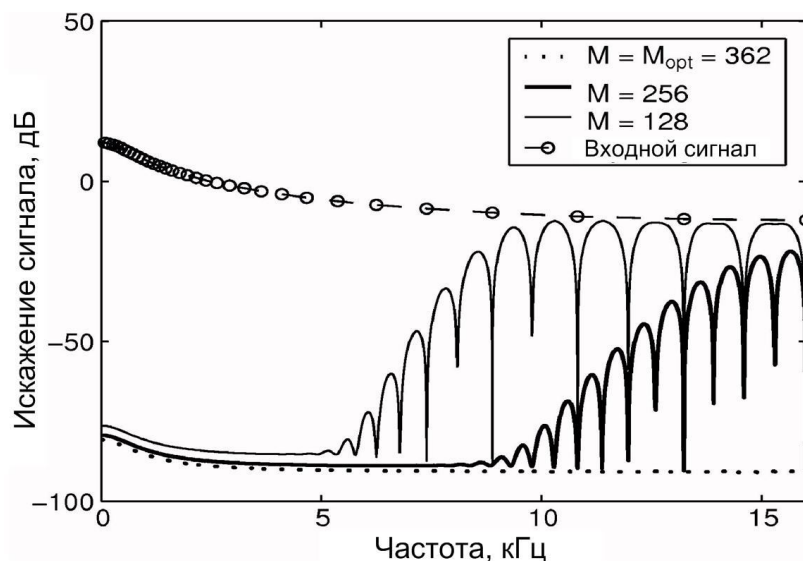


Рис. 14.2 – Спектральная плотность мощности входного сигнала и соответствующие спектральные искажения для разных размеров WDFT преобразования ($N = 64$)

Ниже приведен листинг – программы формирования WDFТ-матрицы и вычисления псевдообратной WDFТ-матрицы на основе разложения по сингулярным числам матрицы.

```
function [W, W_1] = wdfmtx2(N,M,a,cond)
% generation of WDFТ2 matrix of size NxM with warping coefficient a
%[ e^(-jw[1]) e^(-jw[1]) ... e^(-jw[1]) ] .^ [ 0 1 ... M-1 ]
%[ e^(-jw[2]) e^(-jw[2]) ... e^(-jw[2]) ] [ 0 1 ... M-1 ]
%[ ... ... ... ... ] [ ... ... ... ... ]
%[ e^(-jw[N]) e^(-jw[N]) ... e^(-jw[N]) ] [ 0 1 ... M-1 ]
%
% z_1 = exp(-j * 2 * pi * (0 : (N - 1)) ./ N).';
% z_1 = (z_1 - a) ./ (1 - a .* z_1);
%
% W = (z_1*ones(1,M)).^(ones(N,1)*[0:M-1]);
% return;

N2 = ceil(N/2);
z_1 = exp(-j * 2 * pi * (0 : (N2 - 1)) ./ N).';
z_1 = (z_1 - a) ./ (1 - a .* z_1);
W = (z_1*ones(1,M)).^(ones(N2,1)*[0:M-1]);
if N/2 == N2,
    W = [ W; (-ones(1,M)).^[0:M-1]; conj(flipud(W(2:N2,:))) ];
else
    W = [ W; conj(flipud(W(2:N2,:))) ];
end;
if nargin < 2, return; end;

% generation of pseudoinverse matrix via damped SVD
% W = U*S*V', where U*U' = I and V*V' = I
%
% W_1 = V*S_1*U'
%
% S = diag( [ s[1]^(-1) s[2]^(-1) ... s[K]^(-1) 0 ... 0 ] ),
% where s[K] > lambda
[U, S, V] = svd(W, 'econ');
% perform regularization
s = diag(S);
if nargin < 4, cond = 10; end;
lambda = max(s)/cond;
r = sum(s > lambda);
S_1 = diag([ones(r,1)./s(1:r); 0.0*ones(length(s)-r,1)]);
W_1 = V*S_1*U';
```

14.2 Новая схема обработки сигналов в области коэффициентов WDFТ

Таким образом, на основании предложенного выше дискретного преобразования WDFТ можно предложить новую схему обработки речевых и звуковых сигналов в области коэффициентов WDFТ для мультимедиа приложений. Структуры большинства перцептуальных систем редактирования шумов в речи и компрессии звуковых сигналов могут быть представлены обобщенной схемой (см. рис. 14.3) [5,6].

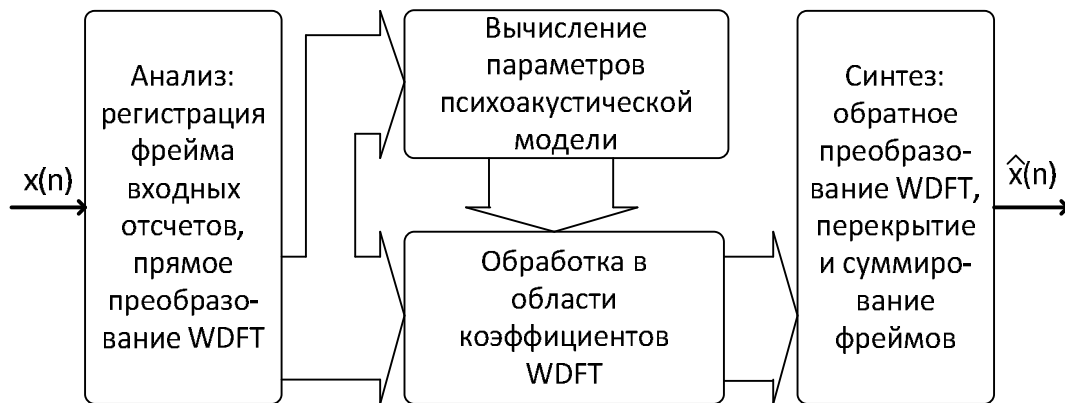


Рис. 14.3 – Схема обработки сигналов в области коэффициентов WDFFT

Входной сигнал $x(n)$ разбивается на квазистационарные фреймы длительностью от 2 до 50 миллисекунд в зависимости от мультимедиа приложения. В блоке анализа осуществляется регистрация фрейма входного сигнала и вычисляется прямое преобразование WDFFT с учетом шкалы критических частотных полос. Далее оценивается энергия маскирующего сигнала (уровни маскирующих порогов) на базе оцененных параметров психоакустической модели. При этом, в процессе обработки в области коэффициентов WDFFT, определяются максимальные искажения, возникающие в каждой точке частотно-временной плоскости без введения искусственного артефакта слышимости в восстановленном сигнале. Следовательно, психоакустический анализатор вычисляет частотно-временной параметр невосприятия акустической информации слушателем. В процессе перцептуальной обработки сигнала необходимо: во-первых, установить вид маскирующего сигнала, во-вторых, вычислить соответствующие пороги, затем полученную информацию использовать для того, чтобы расположить спектр шума обработки ниже порога маскирования, известный как порог едва различимых искажений *JND* (just noticeable distortion) [7]. Таким образом, та часть сигнала, которая может быть изменена (в общем случае отброшена) и при этом не вносится дополнительных искажений при его восстановлении, является перцептуально избыточной, а та часть сигнала, отражающая слышимую акустическую информацию человеком, измеряется и обрабатывается. Выбор же метода частотно-временного анализа неразрывно взаимосвязан с точностью вычисления психоакустической модели и общей концепцией построения системы обработки речевых и звуковых сигналов. При этом выбор методологии частотно-временного анализа всегда обуславливает решение фундаментального вопроса: обеспечения требуемого разрешения как в частотной так и временной областях.

14.3 Схема подавления шума в речевом сигнале на основе спектрального взвешивания

14.3.1 Постановка задачи

Большинство существующих систем подавления шума работают в частотной области, при этом используется хорошо известный подход спектрального вычитания или другими словами правило спектрального взвешивания (рис. 14.4) [8,9].



Рис. 14.4 – Схема подавления шума на основе спектрального взвешивания

Предполагается, что чистый речевой сигнал $s(n)$ и окружающий шум $r(n)$ статистически независимы и стационарны в широком смысле (n обозначает временной индекс). Тогда оцифрованный зашумленный речевой сигнал может быть представлен следующим образом:

$$y(n) = s(n) + r(n). \quad (14.11)$$

Обработка осуществляется в частотной области последовательными фреймами. Входной сигнал $y(n)$ разбивается на перекрывающиеся фреймы из N последовательных отсчетов. Каждый фрейм умножается на оконную функцию и преобразуется в частотную область $Y(\omega)$, где $0 < \omega < 2\pi$. Уменьшение шума достигается умножением спектральных коэффициентов зашумленной речи на действительные коэффициенты взвешивающей функции $H(\omega)$

$$\tilde{S}(\omega) = H(\omega)Y(\omega), 0 \leq H(\omega) \leq 1. \quad (14.12)$$

Следует заметить, что фаза сигнала не изменяется. После модификации амплитуд, спектр реконструированного системой сигнала $\tilde{S}(\omega)$ преобразуется обратно во временную область.

Несмотря на то, что эти методы очень просты и легко реализуются, их слабым местом является остаточный шум, также известный как «музыкальный тон» (см. рис. 14.5 и рис. 14.6). Трехмерная картинка (рис. 14.6) распределения мощности остаточного шума, соответствующая заштрихованной части спектрограммы (рис. 14.5), очень наглядно демонстрирует недостаток данного подхода подавления шума в речевом сигнале.

Следовательно, возникает задача модификации взвешивающего правила на основе принципов психоакустики таким образом, чтобы оставить музыкальные тона немного ниже порога маскирования. Уже достаточно много предложено решений повышения качества речевого сигнала на основе психоакустических моделей [10,11].

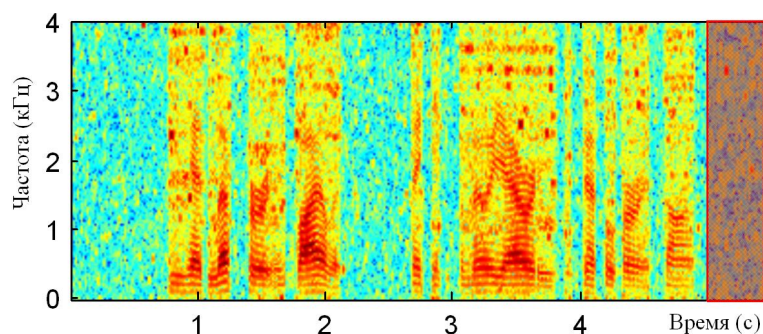


Рис. 14.5 – Спектрограмма выходного сигнала системы подавления шума

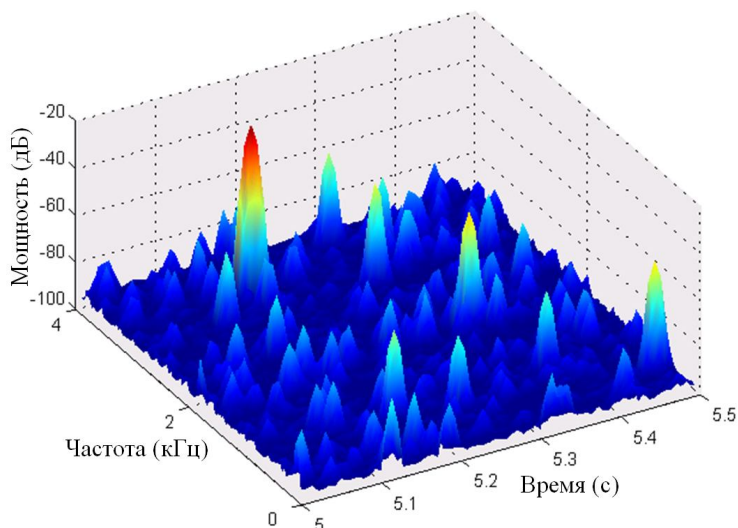


Рис. 14.6 – Трехмерная картинка распределения мощности остаточного шума, соответствующая заштрихованной части спектрограммы на рис. 14.5 - «музыкальный тон»

14.3.2 Перцептуальная система подавления шума в области коэффициентов WDFT

Основная схема подавления шума на основе WDFT показана на рис. 14.7. В данной работе используется наиболее общее психоакустически мотивированное правило спектрального взвешивания [12].

Основной задачей данного подхода является не полное удаление шума, а ослабление слышимого шума до уровня комфортного шума [12]. Это обусловлено тем, что в некоторых случаях, например, во время диалога по телефону, полное удаление шума не желательно. Для того чтобы сохранить характеристики окружающего шума необходимо определить предпочтительный уровень остаточного шума ζ_n . Тогда разница между желаемым спектром очищенной речи и его оценкой может быть определена как

$$Q(\omega) = S(\omega) + \zeta_n S(\omega) - H(\omega)[S(\omega) + N(\omega)], \quad (14.13)$$

где $S(\omega)$ и $N(\omega)$ спектры чистой речи и шума, соответственно, и $H(\omega)$ – оценка весовой функции.

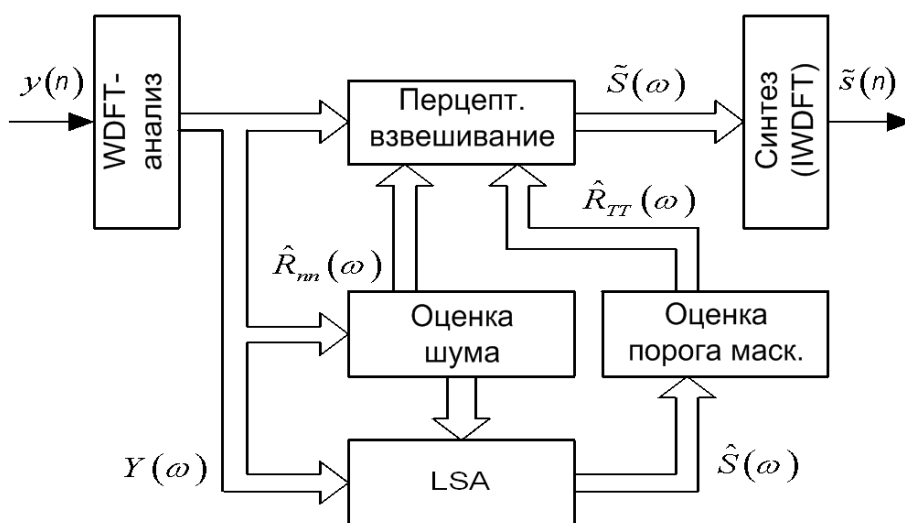


Рис. 14.7 Структура перцептуальной системы подавления шума на базе WDFT

Так как речь и шум предполагаются статистически независимыми, то спектральная плотность мощности (СПМ) разности (14.13) может быть выражена следующим образом:

$$R_{qq}(\omega) = [1 - H(\omega)]^2 R_{ss}(\omega) + [\zeta_n - H(\omega)]^2 R_{nn}(\omega), \quad (14.14)$$

или

$$R_{qq}(\omega) = R_{q_s q_s}(\omega) + R_{q_n q_n}(\omega). \quad (14.15)$$

Слагаемые выражения (14.15) соответствуют СПМ искажения речи и шума, соответственно. Для минимизации суммы (14.15) в перцептуальном смысле необходимо, чтобы уровень искажений был не воспринимаемым на слух. В идеальном случае все искажения должны быть замаскированы. Тем не менее, в большинстве реальных систем это требование не может быть удовлетворено, так как минимум $R_{qq}(\omega)$ может быть больше, чем СПМ порога маскирования $R_{TT}(\omega)$. Поэтому критерий минимизации формулируется как

$$R_{q_n q_n}(\omega) = R_{TT}(\omega). \quad (14.16)$$

Решая уравнение (14.16) относительно $H(\omega)$, определяется весовая функция

$$H^{IND}(\omega) = \min \left\{ 1, \sqrt{\frac{R_{TT}(\omega)}{R_{nn}(\omega)} + \zeta_n} \right\}, \quad (14.17)$$

где IND означает неслышимое шумовое искажение. Легко заметить, что если остаточный шум лежит ниже порога маскирования, выражение под корнем больше единицы и речь не искажается так как $H^{IND}(\omega) = 1$. В противном случае, окружающий шум оптимально ослабляется до уровня, не воспринимаемого на слух человеком. Заметим, что оценки порога маскирования и СПМ шума необходимы только для вычисления взвешивающих коэффициентов. Простая модификация перцептуальной энтропии [13] используется как базис для оценки порога маскирования чистой речи. Пре-процессор оценки СПМ зашумленного речевого сигнала реализован на основе метода LSA [9] (рис. 14.7).

Применение WDFT в перцептуальной системе редактирования шума не только как базиса для определения модели маскирования, но и как инструмента спектрального анализа, позволяет добиться лучшего качества реконструированного речевого сигнала по сравнению с системами на основе ДПФ. Это объясняется тем, что весь процесс обработки осуществляется в перцептуальном домене с неравномерной частотной шкалой и нет необходимости в преобразованиях между разными частотными шкалами, что приводит к упрощению архитектуры системы. Более того, обработка речи, осуществляемая в критических частотных полосах, более точна в контексте психоакустического моделирования.

14.4 Оценка порога маскирования и спектральной плотности мощности шума

14.4.1 Отображение коэффициентов WDFT на критические частотные полосы

Исследования, проведенные в работах [4,14], показали, что малоформатное WDFT может успешно заменить ДПФ с большой длиной выборки. Это оказалось возможным благодаря тому, что WDFT позволяет разместить частотные компоненты в соответствии с распределением критических частотных полос, поэтому в психоакустической модели на базе WDFT могут быть уравновешены как хорошее частотное, так и временное разрешение [15].

На основе свойств WDFT процедура оценки порога маскирования на базе общей психоакустической модели Джонстона [13] была модифицирована. Первый шаг при использовании WDFT в психоакустической модели – проектирование соответствующего фазового звена [4,15]. Частотные коэффициенты z - преобразования должны быть представлены регулярно в перцептуальной области. Показано, что фазовое звено первого порядка достаточно хорошо аппроксимирует перцептуальную шкалу барков, При этом значение коэффициента фазового звена для частоты дискретизации f_s определяется по следующему выражению [15]:

$$a_{bark} = 0.1957 - 1.048 \cdot \left[\frac{2}{\pi} \arctan \left(0.07212 \frac{f_s}{1000} \right) \right]^{\frac{1}{2}}. \quad (14.18)$$

Табл. 14.1 – Распределение коэффициентов ДПФ и WDFT в критических частотных полосах

Крит. полоса	Часть А (размер ДПФ = 256, $f_s=8$ кГц)			Часть Б (размер WDFT = 256, $f_s=8$ кГц)		
	Диапазон коэффициентов	Кол-во	Диапазон частот, Гц	Диапазон коэффициентов	Кол-во	Диапазон частот, Гц
1	1 – 3	3	31 - 94	1 – 7	7	13 – 92
2	4 – 6	3	125 - 188	8 – 15	8	105 – 198
3	7 – 9	3	219 - 281	16 – 22	7	212 – 294
4	10 – 12	3	313 - 375	23 – 29	7	308 – 394
5	13 – 16	4	406 - 500	30 – 36	7	408 – 498
6	17 – 20	4	531 - 625	37 – 44	8	514 – 627
7	21 – 24	4	656 - 750	45 – 52	8	644 – 768
8	25 – 29	5	781 - 906	53 – 59	7	786 – 904
9	30 – 34	5	938 - 1063	60 – 67	8	925 – 1080
10	35 – 40	6	1094 - 1250	68 – 74	7	1103 - 1255
11	41 – 47	7	1281 - 1469	75 – 81	7	1282 - 1458
12	48 – 55	8	1500 - 1719	82 – 88	7	1489 - 1694
13	56 – 64	9	1750 - 2000	89 – 95	7	1731 - 1972
14	65 – 74	10	2031 - 2313	96 – 102	7	2015 - 2301
15	75 – 86	12	2344 - 2688	103 – 109	7	2352 - 2688
16	87 – 100	14	2719 - 3125	110 – 116	7	2748 - 3134
17	101 – 118	18	3156 - 3688	117 – 123	7	3202 - 3629
18	119 – 128	10	3719 - 4000	124 – 128	5	3703 - 4000

Для случая $f_s = 8$ кГц коэффициент $a_{bark} = -0.4092$. Так как ширина критических частотных полос строго изменяется с их местоположением на частотной шкале, то количество коэффициентов преобразования ассоциируется с конкретной критической частотной полосой (табл. 14.1).

WDFT в его оригинальной форме не сохраняет энергию сигнала в соответствующих частях единичной окружности до и после фазового звена. Вследствие этого, каждый коэффициент WDFT должен масштабироваться в соответствии с коэффициентом $\sqrt{1 - a^2}/(1 - az)$ [5], чтобы вычислить корректно уровни мощности в критических частотных полосах.

14.4.2 Оценка СПМ шума

Оценка дисперсии шума является ключевой задачей для многих систем повышения качества речевого сигнала. Наиболее общий подход основывается на статистических измерениях в течение речевых пауз, используя экспоненциальное усреднение. Периоды речевых пауз определяются в зашумленном речевом сигнале на основе детекторов вокализованности речевого фрейма (VAD). Следовательно, эффективность таких анализаторов шума строго зависит от уровня SNR и типа шума [16]. В частности, они экстремально чувствительны к внезапным изменениям уровня шума. Другие, более робастные подходы базируются на методе минимальной статистики [17]. Ключевая идея которого, заключается в слежении за минимальным уровнем мощности шума для каждой спектральной компоненты на протяжении ряда фреймов сигнала. В случае уменьшения мощности шума – быстрая корректировка оценки мощности шума здесь очевидна. Но в противоположной ситуации (мощность шума увеличивается) - обновление результата оценки задерживается на целый период слежения. К сожалению, метод минимальной статистики не совсем очевидный и получение оптимальной компенсационной процедуры достаточно затруднительно [18]. В данной системе (рис. 14.7) используется модифицированная процедура экспоненциального усреднения с контролем по минимуму энергии (MCRA) [19]. Этот подход находится между обычным экспоненциальным усреднением и методом минимальной статистики. Далее в главе показано как оптимизировать данный метод, используя пещептуально минимизированное усреднение в области WDFT коэффициентов.

В оригинальном MCRA алгоритме [19], оценка СПМ зашумленной речи основывается на методе периодограмм. В данном решении оценки СПМ сглаживаются в соответствующих критических частотных полосах – барках. А именно, предполагается, что для большинства практических применений при оценке СПМ зашумленной речи в критических частотных полосах справедливо утверждение [20], т.е.

$$R_{yy}(k, l) \approx \hat{R}_{yy}(b, l), \text{ для } k \in K_b, \quad (14.19)$$

где l - номер обрабатываемого фрейма, b – номер критической частотной полосы; K_b – количество коэффициентов преобразования, попадающих в b -ю критическую частотную полосу; k – номер коэффициента преобразования в b -й критической частотной полосе. Во временной области усреднение осуществляется следующим образом:

$$\hat{R}_{yy}(b, l) = \alpha \hat{R}_{yy}(b, l - 1) + (1 - \alpha) \hat{S}_{yy}(b, l). \quad (14.20)$$

Оценка СПМ зашумленной речи на основе мгновенных оценок СПМ $|Y(k, l)|^2$ в b -й критической частотной полосе равна:

$$\hat{S}_{yy}(b, l) = \frac{1}{\omega_{high,b} - \omega_{low,b}} \sum_{k \in K_b} \Delta\omega_k |Y(k, l)|^2, \quad (14.21)$$

где $\omega_{high,b}$, $\omega_{low,b}$ - границы b -ой критической частотной полосы и $\Delta\omega_k$ полоса k -го фильтра в WDFТ спектре. Заметим, что выражение (14.21) представляет аппроксимацию процесса интегрирования непрерывной функции СПМ по пространству неравномерно расположенных частотных отсчетов. Для стандартного ДПФ с постоянным частотным разрешением, оценка (14.21) представляет собой среднее арифметическое мгновенных значений СПМ $|Y(k, l)|^2$. Сглаживание внутри критических частотных полос уменьшает дисперсию оценки СПМ. Хотя при этом разрешающая способность спектрального анализа также уменьшается. Однако это не скажется на окончательном результате редактирования шума в речевом сигнале, потому что большинство шумов имеют равномерные спектры и более того, разрешающая способность уменьшается в соответствии с частотной избирательностью человеческого уха. Таким образом, подобно (14.19) можно утверждать, что при оценке СПМ шума в каждой критической частотной полосе выполняется равенство:

$$R_{nn}(k, l) \approx \hat{R}_{nn}(b, l), \text{ для } k \in K_b, \quad (14.22)$$

Согласно MCRA схеме, СПМ шума в b -ой критической частотной полосе и для l -ого фрейма сигнала оценивается следующим образом:

$$\hat{R}_{nn}(b, l) = \tilde{\alpha}_n(b, l) \hat{R}_{nn}(b, l-1) + (1 - \tilde{\alpha}_n(b, l)) \hat{S}_{yy}(b, l), \quad (14.23)$$

где $\tilde{\alpha}_n(b, l)$ время зависимый коэффициент сглаживания, являющийся функцией оценки вероятности $p(b, l)$ вокализованности (присутствия речевой активности) l -ого фрейма речевого сигнала:

$$\tilde{\alpha}_n(k, l) = \alpha_n + (1 - \alpha_n)p(b, l), 0 < \alpha_n < 1. \quad (14.24)$$

Коэффициент сглаживания α_n контролирует общую способность слежения при оценке СПМ шума. Вероятность $p(b, l)$ вокализованности l -ого фрейма рассчитывается на основе экспоненциального усреднения решений $I(b, l)$ несложного VAD, реализованного на подходе минимума статистики:

$$p(b, l) = \alpha_p p(b, l-1) + (1 - \alpha_p) I(b, l), 0 < \alpha_p < 1, \quad (14.25)$$

где α_p - коэффициент сглаживания. Решение VAD $I(b, l)$ интерпретируется как бинарный результат сравнения СПМ $\hat{R}_{yy}(b, l)$ зашумленного речевого сигнала и оценки СПМ детектора минимума статистики $R_{min}(b, l)$ в b -ой критической частотной полосе l -ого фрейма сигнала:

$$I(b, l) = \begin{cases} 1, & \text{если } \hat{R}_{yy}(b, l) \geq B \cdot R_{min}(b, l) \\ 0, & \text{иначе} \end{cases}. \quad (14.26)$$

Здесь параметр B выбирается эмпирически как положительная константа (обычно $2 < B < 3$). Оценка детектора минимума статистики получается согласно правилу:

$$R_{min}(b, l) = \min\{\hat{R}_{yy}(b, l - T), \hat{R}_{yy}(b, l - T + 1), \dots, \hat{R}_{yy}(b, l)\}. \quad (14.27)$$

где T обозначает длину окна поиска минимума.

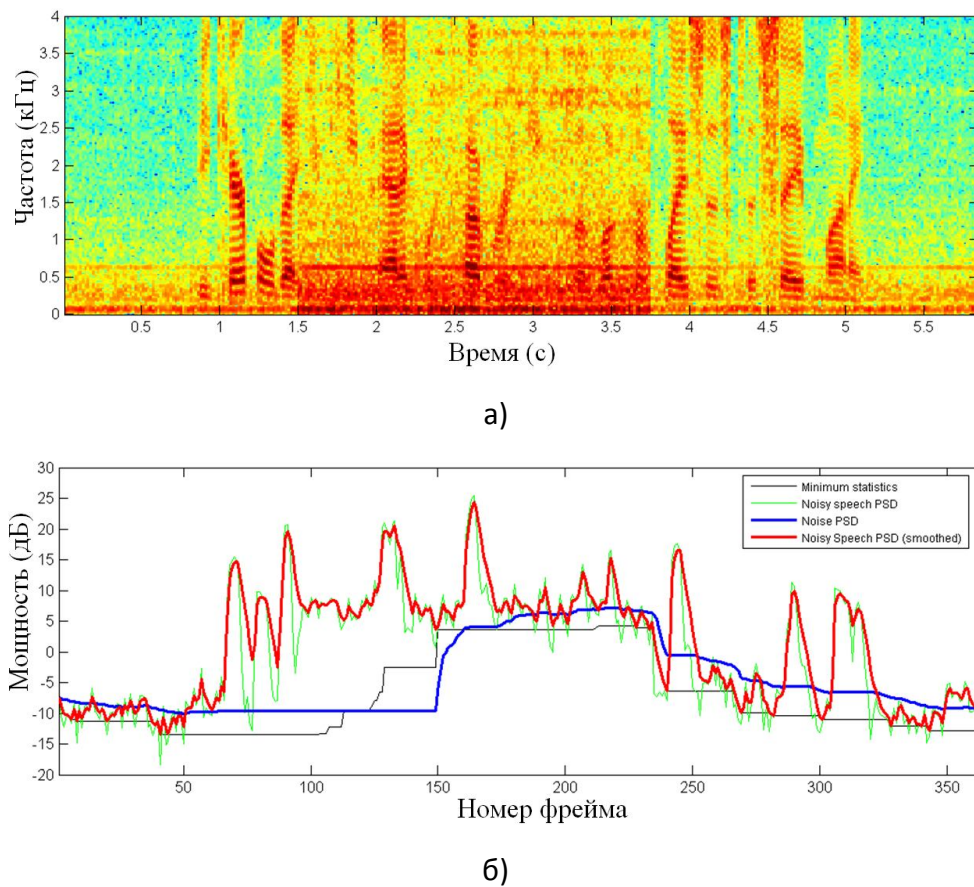


Рис. 14.8 – Пример слежения за изменением СПМ на основе алгоритма MCRA: а) спектрограмма зашумленного речевого сигнала; б) изменения СПМ

На рис. 14.8 показан пример слежения в зашумленном речевом сигнале за изменением СПМ, у которого внезапно изменился уровень мощности шума. В данном эксперименте параметр $T = 60$ фреймам. Как можно видеть из рис. 14.8б отклик алгоритма MCRA на скачок мощности шума имеет заметное опоздание за счет длины окна поиска минимума. В случае уменьшения уровня мощности, как показывает эксперимент, алгоритм реагирует незамедлительно. На рис. 14.8а представлена спектрограмма данного зашумленного речевого сигнала.

14.4.3 Оценка порога маскирования

Основные шаги оценки порога маскирования следующие: расчет энергии критических частотных полос по WDFT-спектру мощности; свертка с функцией распространения; вычитание относительного смещения порога; нормализация и сравнение с абсолютным порогом слышимости [21,22].

Дискретное представление энергии критической частотной полосы может быть выражено как взвешенная сумма компонент спектра мощности:

$$E_b = \frac{1}{2\pi} \sum_{k \in K_b} \Delta\omega_k |\hat{S}(\omega_k)|^2, \quad (14.28)$$

где b – номер критической частотной полосы; K_b – количество коэффициентов преобразования, попадающих в b -ю критическую частотную полосу; k – индекс коэффициента преобразования (спектрального отсчета); $\hat{S}(\omega_k)$ – оценка спектра оригинального речевого сигнала.

На следующем шаге вычисляется свертка энергий критических частотных полос E_b с функцией распространения по базилярной мембране SF_b для учета распространения маскирования:

$$C_b = E_b \cdot SF_b. \quad (14.29)$$

Относительное смещение порога в каждой критической частотной полосе рассчитывается с использованием меры тональности. Для определения типа маскера («шумоподобный» или «тональный») используется мера пологости спектра, которая определяется как отношение среднего геометрического μ_{gb} к среднему арифметическому μ_{ab} спектральных компонент внутри критической полосы b :

$$SFM_b = 10 \log_{10} \left(\frac{\mu_{gb}}{\mu_{ab}} \right), \text{ [дБ]}. \quad (14.30)$$

Получение данной оценки может быть затруднено в случае обычных моделей с малым разрешением, так как количества коэффициентов ДПФ в низкочастотных полосах обычно не достаточно для эффективного расчета SFM . В подобных случаях SFM рассчитывается для всего спектра или predetermined тональностей для каждой критической частотной полосы. В случае WDFT, коэффициенты преобразования группируются равномерно в критических частотных полосах (табл. 14.1), следовательно, SFM может быть определена для каждой полосы отдельно.

Смещение порога маскирования O_b оценивается по следующему выражению:

$$O_b = ton_b \cdot (14.5 + b) + (1 - ton_b) \cdot 5.5, \text{ [дБ]}, \quad (14.31)$$

где ton_b – индекс тональности, определяемый как

$$ton_b = \min \left(\frac{SFM_b}{-60 \text{ дБ}}, 1 \right). \quad (14.32)$$

Для получения энергии порога маскирования относительное смещение (14.31) вычитается из свернутого спектра критической полосы (14.29):

$$P_{TT,b} = 10^{\left(\log_{10} C_b - \frac{O_b}{10} \right)}, \quad (14.33)$$

а затем осуществляется нормализация и сравнение с абсолютным порогом слышимости.

Таким образом, оценки СПМ шума, также как и оценки порога маскирования, вычисленные на их основе, на обрабатываемом фрейме речевого сигнала постоянны в критических частотных полосах, а следовательно, и соответствующие спектральные взвешивающие коэффициенты внутри критических частотных полос также постоянны. Это свойство не только упрощает схему обработки, но также ослабляет генерацию «музыкальных тонов».

Ниже приведена MATLAB-программа вычисления порогов маскирования в области коэффициентов WDFT.

```

%WDFT_MT Masking Threshold Estimate
% [Ptt, Pqq] = mt(Pxx, freqs) estimates masking threshold
%                                     for the current signal frame
% input:  Pxx  - PSD of the current signal frame
%         freqs - frequencies of the fft bins
% output: Ptt  - PSD estimate of the masking threshold
%         Pqq  - Absolute threshold of hearing

function varargout = wdft_mt(Pxx, freqs)

% Critical bandwidth calculation (in dft bins)
BW  = cb_bins(freqs);
n_BW = length(BW);

% Calculation (SPL calibrating) of absolute threshold of hearing
% |X| = A*N/2 => Pxx = |X|^2 = A^2 * N^2 / 4
% for 16bit sampling we have
% Pxx(Amin) = (2^-15)^2*N^2/4 < 2^-15 => Pxx(Amin) = Amin

Pxx_min = 10*log10(2^-12); % can be fixed experimentally too
Pqq = tq(freqs);
Pqq = 10.^((Pqq - min(Pqq) + Pxx_min)/10);

% Spreading function calculation (10-Bark masking range)
SF  = 10.^(spreadf(-3:6)/10);
n_SF = length(SF);

% Conversion of Pxx to Bark domain
fs  = 2*freqs(end);
Bxx = pw2bark(Pxx, freqs, fs, BW);

% Convolution with spreading function
D = conv(Bxx, SF);
D = D(floor((n_SF-1)/2):floor((n_SF-1)/2)+n_BW-1);

% Offset calculation
a = zeros(n_BW, 1);
b = 1;
for k = 1:n_BW
    a(k) = min(sfm(Pxx(b:b+BW(k)-1))/-60, 1);
    % (noisy) 0 < a < 1 (tonal)
    b = b + BW(k);
end;
Off = (a).*(14.5+(1:n_BW)')+ (1-a).*5.5; % [dB]

% Masking threshold
Btt = 10.^(log10(max(D, 1e-6)) - Off/10); % [power]

% Conversion to original scale

```

```

Ptt = bark2pw(Btt, freqs, fs, BW);

% Comparison with ATH
Ptt = max(Ptt, Pqq);

if nargout>0, varargout{1} = Ptt; end;
if nargout>1, varargout{2} = Pqq; end;
return;

%CB_BINS - groups dft bins into critical bands
% input:  freqs - dft bins frequencies
%         method - uses predefined set
% output: BW    - amount of dft bins in critical bands

function BW = cb_bins(freqs, method)
% Critical bandwidth calculation (single Bark resolution)

if nargin == 1, % auto
    BW = zeros(floor(1+bark(max(freqs))),1);
    npsd = length(freqs);
    for k = 1:npsd,
        b = floor(1+bark(freqs(k)));
        BW(b) = BW(b) + 1;
    end;
else % manual
    switch lower(method)
    case 'dft8khz'
        BW = [3 3 3 3 4 4 5 5 5 6 6 8 8 11 13 16 20 6]';
    case 'wdft8khz'
        BW = [8 8 8 8 7 7 7 7 7 7 7 7 7 7 7 7 6]';
    end;
end;
return;

%Absolute threshold of hearing [dB SPL]
function y = tq(f)

y = 3.64*(f/1000).^(-0.8)-6.5*exp(-0.6*(f/1000-3.3).^2)+0.001*(f/1000).^4;
return;

% Spreading function [dB]
function y = spreadf(x)
y = 15.81 + 7.5*(x + 0.474) - 17.5*sqrt(1 + (x + 0.474).^2);
return;

%PW2BARK calculates normalized subband energies for warped DFT
% input:
% Sxx = |X|^2 - WDFFT power spectrum (one sided)
% freqs - vector with warped center frequencies for DFT bins
% fs - frequency sampling
% BW - DFT bins grouping

```

```

% output
% Bxx - vector of subband energies, note that sum(Bxx) = sum(x.^2)
function Bxx = pw2bark(Sxx, freqs, fs, BW)

%          fu(k)
% Bxx(k) = 1/fs * S    |X(f)|^2 df    - Band energy
%          fl(k)

n_BW = length(BW);
Bxx = zeros(n_BW, 1);
b = 1;
for k = 1:n_BW
    for j = b:b+BW(k)-1
        df = (freqs(min(length(freqs), j+1)) - freqs(max(1, j-1))) / 2;
        Bxx(k) = Bxx(k) + Sxx(j) * df / fs;
    end;
    b = b + BW(k);
end;
return;

```

```

% Conversion of Bxx from Bark scale to user-defined scale
function Sxx = bark2pw(Bxx, freqs, fs, BW)

n_BW = length(BW);
b = 1;
for k = 1:n_BW
    df = (freqs(min(length(freqs), b+BW(k))) + freqs(b+BW(k)-1) - ...
          freqs(b) - freqs(max(1, b-1))) / 2;
    Sxx(b:b+BW(k)-1, 1) = Bxx(k) * fs / df;
    b = b + BW(k);
end;
return;

```

```

% Spectral Flatness Measure [dB]
function m = sfm(X)

eps = 1e-6;
X = max(X, eps);
m = 10 * log10(max(geomean(X) / max(mean(X), eps), eps));
return;

```

```

% Converts linear frequency in Hz to Bark scale
function b = bark(f)
b = 13 * atan(0.00076 * f) + 3.5 * atan((f / 7500).^2);
return;

```

14.5 Оценка качества системы подавления шума

Эксперименты по оценке качества реконструированного системой речевого сигнала проводились для следующих параметров настройки алгоритмов: частота дискретизации вход-

ного сигнала – 8 кГц, размер фрейма $N = 256$ отсчетов (32 мс), обработка осуществляется с 50% перекрытием фреймов, временное окно – Хеннинга. В порядке сокращения вычислительной сложности, размер окна анализа WDFT с избыточным базисом [15] был сокращен до 128 отсчетов. Как было показано в [4], оптимальный формат WDFT для $N = 128$ и коэффициента фазового звена $a_{Bark} \approx -0.4$ находится около 300. Данную величину можно немного уменьшить благодаря перцептуальному маскированию искажений в речевом сигнале. Таким образом, размер матрицы $\mathbf{D}_{M \times N}$ WDFT с избыточным базисом был выбран 256×128 .

MATLAB-программа редактирования шума окружающей среды в речевом сигнале в области коэффициентов WDFT представлена ниже.

```

% WDFT based speech enhancement system
% S = wdft_nr(X, FS) estimates clean speech signal
% input:
%     X - time sampled noisy speech signal
%     FS - sampling frequency
% output:
%     S - enh. speech signal
function varargout = wdft_nr(x, fs)

%[x, fs] = wavread('SI805.noi.wav');

x = x(:);           % input signal
s = zeros(size(x)); % output signal
nw = 256;           % frame size
nw_2 = floor(nw/2)+1;
no = 128;           % overlapping 50%
W = hann(nw);       % analysis window

% Computation of the WDFT matrices
a = 0.1957 - 1.0480 * sqrt((2 / pi) * atan(0.00007212 * fs));
[F, F_1] = wdftmtx2(nw, nw, a);
freqs = fs/(2*pi) * abs(angle(F(1:nw_2,2)));
alpha = 0.2; % noisy speech periodogram smoothing
theta = 0.5; % noise periodogram smoothing
beta = 0.98; % LSA smoothing parameter

% VAD threshold
thr = 1.5 * log10(x(1:nw)'*x(1:nw));
for n = 1:nw-no:length(x)-nw,
    % Analysis
    select = n:n+nw-1;
    X = F * (x(select) .* W);

    % Periodogram
    Pxx = abs(X(1:nw_2)).^2;
    if n==1,
        Rx = Pxx;
        Rn = Pxx;
        Sxx = Pxx;
    end;
end;

```

```

% Smoothed periodogram (time-frequency averaging)
Rx = Rx.*(1-alpha) + Pxx.*alpha;

% Update noise estimate (log-energy VAD)
if log10(x(select)'*x(select)+eps) < thr,
    Rn = Rn.*(1-theta) + Pxx.*theta;
end;

% LSA rule
SNRpost = max(Pxx./Rn, 1+1e-6);
SNRprio = max(beta.*(Sxx./Rn) + (1-beta).*(SNRpost-1), 1e-6);
Gw      = SNRprio./(SNRprio + 1);           % Wiener rule
Glsa    = Gw.*exp(0.5*expint(Gw.*SNRpost)); % LSA rule
S        = X(1:nw_2).*min(Glsa,1);

% Periodogram of the enhanced speech
Pxx = abs(S(1:nw_2)).^2;
Sxx = Pxx;

% Masking threshold est.
Rt = wdft_mt(Pxx, freqs);

% Perceptual Weighting
Gind = min((Rt./(Rn+eps)).^0.5 + 0.05, 1); % IND rule
S      = X(1:nw_2).*Gind;

% Synthesis
S(nw_2+1:nw) = flipud(conj(S(2:nw_2-1+rem(nw,2))));
s(select) = s(select) + real(F_1*S)./max(W);
end;
if nargout > 0, varargout{1} = s; end;
if nargout > 1, varargout{2} = x; end;
%subplot(2,1,1);
%    specgram_abs(x,fs);
%subplot(2,1,2);
%    specgram_abs(s,fs);

% SPECGRAM_ABS - calculates spectrogram from signal w/o prescaling
% [Px, f, t] = SPECGRAM_ABS(x, fs, DRAW)
% x      - signal vector
% fs     - frequency sampling rate (Hz), default is 8000 Hz
% DRAW  - drawing mode ('SCREEN', 'PRINTER', 'NONE'), default is 'SCREEN'
% Px     - matrix with short-time power spectra placed in columns
%                                     (power units)
% f      - frequencies vector (kHz)
% t      - time vector (seconds)
function varargout = specgram_abs(x, fs, draw)
error(nargchk(1, 3, nargin));
if (nargin < 2) fs = 8000; end;
if (nargin < 3) draw = 'SCREEN'; end;
x = x(:);
Nx = length(x);

```

```

nfft = 512; % fft size with zero-padding
Nw = 256; % analysis window
npsd = floor(nfft/2)+1; % size of one-sided psd
No = floor(0.5*Nw); % number of overlapped samples
w = hann(Nw); % window function
w = w*sqrt(Nw/sum(w.^2)); % normalization
L = floor((Nx-Nw)/(Nw-No)); % est. number of frames
Px = zeros(npsd,L);
sel = 1:Nw;

% short-time fourier analysis
for l = 1:L,
    X = fft(x(sel).*w, nfft);
    X = X(1:npsd);
    Px(:,l) = X.*conj(X) ./ Nw;
    sel = sel + (Nw - No);
end;

% labels for time/frequency axes
f = (fs/1000)*(0:npsd-1)'/nfft; % (kHz)
t = (Nw-No)*(1:L)'/fs; % (s)

% spectrogram drawing
if (strcmp(upper(draw),'NONE') ~= 1)
    imagesc(t, f, 10*log10( Px + eps ), [-100 0]), axis xy;
    %image(t, f, 64*(10*log10( Px + eps )+90)/90), axis xy;
    xlabel('Time (s)');
    ylabel('Frequency [kHz]');
    switch (upper(draw))
        case 'SCREEN',
            colormap(jet);
        case 'PRINTER'
            colormap(flipud(gray));
        otherwise
            error('Unknown drawing mode');
    end;
end;

%figure;
%surf(10*log10( Px(150:end,200:end) + eps ));

% optional output arguments
if nargout>0, varargout{1} = Px; end;
if nargout>1, varargout{2} = f; end;
if nargout>2, varargout{3} = t; end;

```

Тестирование программы подавления шума осуществлялось на 8 озвученных предложениях русского языка. Длина речевых сегментов колебалась от 5 с до 8 с. К каждому сегменту чистого речевого сигнала добавлялся «цветной» шум. Сегментное отношение сигнал – шум (SNR_{SEG}) варьировалось от -5 дБ до 20 дБ. Использовались следующие объективные показатели качества: «Искажение речи» (в дБ), «Кепстральное расстояние», «Ослабление шума» в дБ.

Оценка «Искажение речи» (в дБ) это SNR_{SEG} , где под шумом понимается разница между оригинальным (чистым) речевым сигналом и выходным сигналом системы подавления шума. Высокое значение данного SNR_{SEG} показывает на малые искажения речи. Большое значение показателя «Кепстральное расстояние» означает о наличии в выходном сигнале системы сильных артефактов речи. Под оценкой «Ослабление шума» понимается отношение мощности входного зашумленного речевого сигнала к мощности выходного сигнала системы, в котором шум был ослаблен. Однако объективные показатели качества слабо коррелируют с результатами оценок на основе восприятия человеком речи. Поэтому при тестировании использовался показатель «Перцептуальные искажения» на основе модифицированного метода искажений спектра барков (от англ. Modified Bark Spectral Distortion (MBSD)) [23], который определяется как разница между восприятием чистой речи и синтезированной системой.

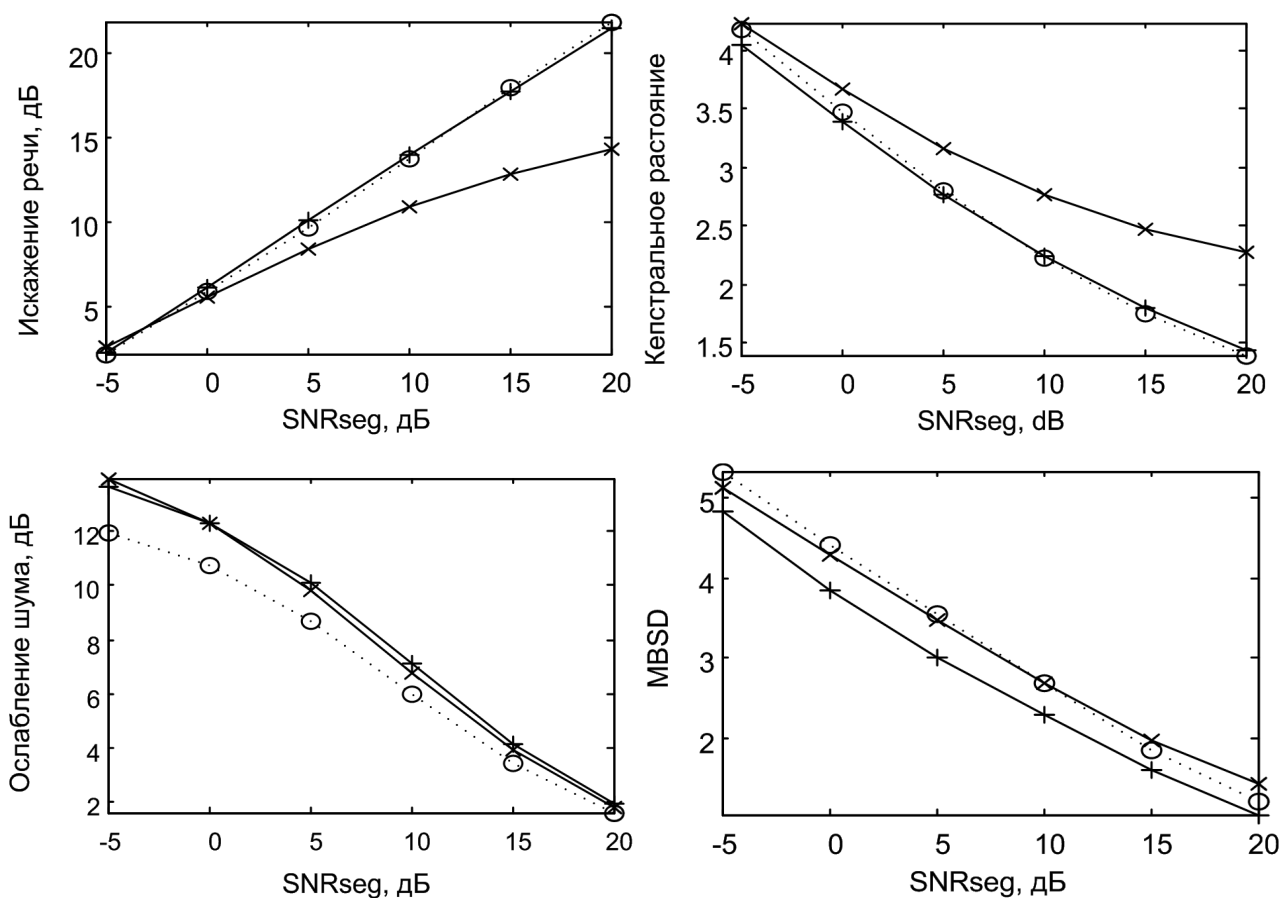


Рис. 14.9 – Оценка качества систем на базе ДПФ (+), WDFT (x), WDFT с избыточным базисом (o)

Экспериментальные результаты проиллюстрированы на рис. 14.9 для системы подавления шума на основе стандартного ДПФ, WDFT «чистого» и с избыточным базисом. Результаты для WDFT с избыточным базисом значительно лучше для оценки «Ослабление шума», чем для ДПФ и «чистого» WDFT. Оценки показателей «Искажение речи» и «Кепстральное расстояние» системы на базе ДПФ и WDFT с избыточным базисом приблизительно одинаковые и много лучше, чем для «чистого» WDFT. Это обусловлено значительным уменьшением ошибки синтеза у алгоритма WDFT с избыточным базисом, по сравнению с «чистым» WDFT [15]. Анализ показателя MBSD показывает, что качество системы на основе WDFT с избыточным базисом не на много выше, по сравнению с системами подавления шума на базе ДПФ и «чистого» WDFT. Это

подтверждает предположение, что генерируемые искажения системой синтеза на основе «чистого» WDFT только отчасти слышны.

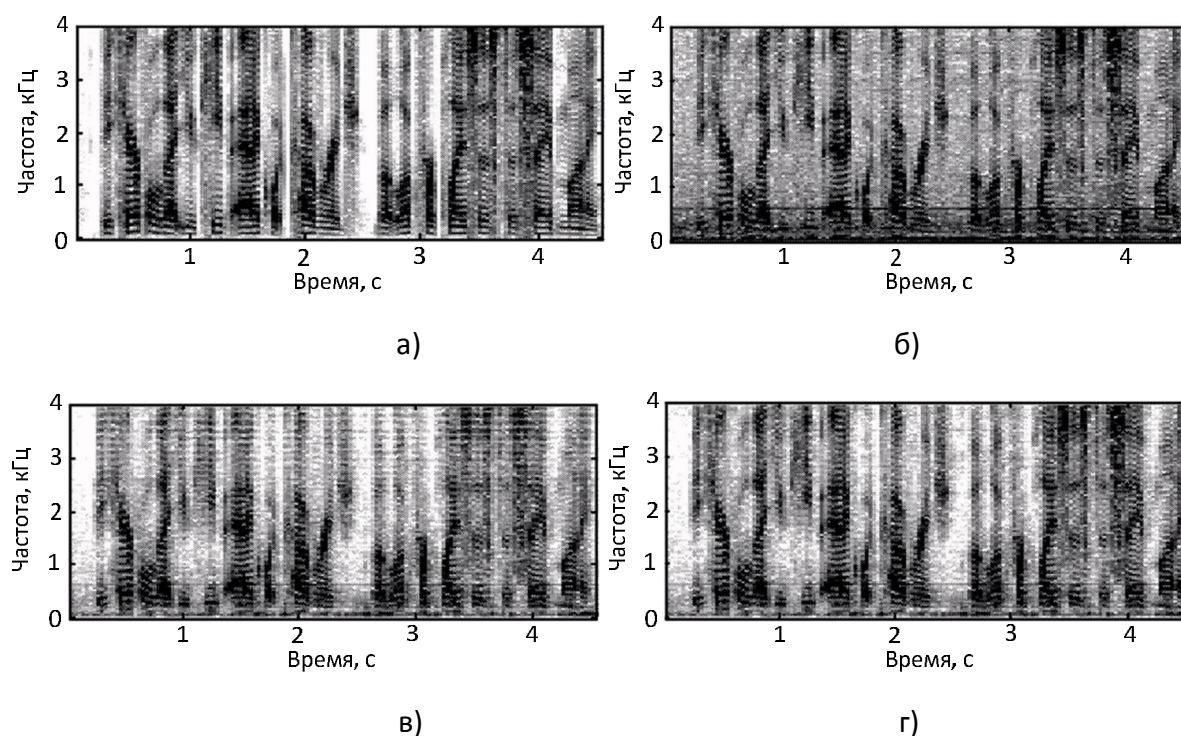


Рис. 14.10 Спектрограммы: оригинальный сигнал (а); зашумленный речевой сигнал $SNR_{SEG} = 5$ дБ (б); сигнал, очищенный системой на базе WDFT (в); сигнал, очищенный системой на базе WDFT с избыточным базисом (г)

Для анализа распределения мощности остаточного шума и искажений речи по частотному диапазону использовались спектрограммы (рис. 14.10). Как видно, нет заметной разницы между системами в низкочастотном диапазоне. Во всех случаях шум окружающей среды сильно ослабляется и «музыкальные» тона не воспринимаются благодаря относительно высокому предопределенному уровню остаточного шума (14.13). Однако, для системы на основе WDFT с избыточным базисом, в отличие от решения с «чистым» WDFT, ясно видно, что высокочастотные искажения не генерируются и высшие гармоники в реконструированном системой речевом сигнале близки к оригинальному сигналу.

Дальнейшие работы в данном направлении нацелены на уменьшение вычислительной сложности и разработку WDFT с высоким частотным разрешением на основе психоакустической моделей (возможно для ERB модели).

14.6 Компрессия звуковых сигналов в области коэффициентов WDFT

14.6.1 Структура системы компрессии звуковых сигналов

Согласно новой схеме обработки сигналов в области коэффициентов WDFT (рис. 14.3) структура системы компрессии звуковых сигналов может быть представлена следующим образом (рис. 14.11) [4,6,24]. Основными компонентами кодера является прямое преобразование

WDFT, блок оценки порогов маскирования, блоки квантования и кодирования с минимальными потерями на основе кодовых книг Хаффмана.

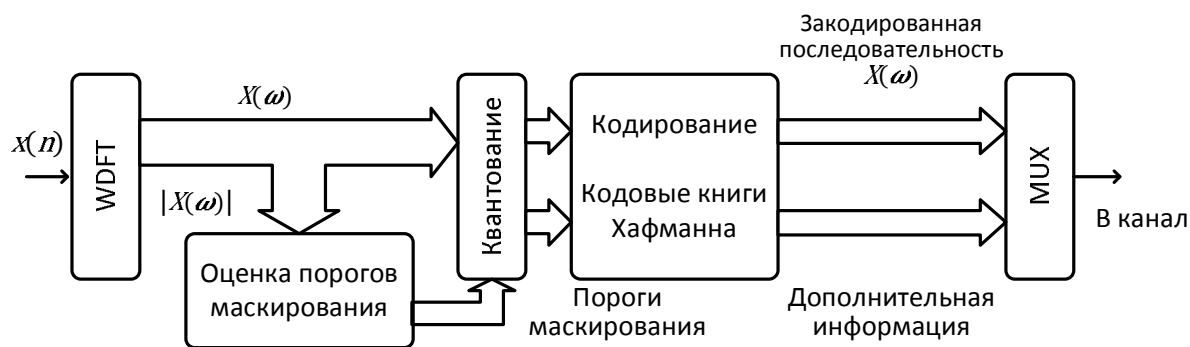


Рис. 14.11 – Структурная схема широкополосного перцептуального кодера речи и звука на основе WDFT

Табл. 14.2 – Распределение коэффициентов WDFT в критических частотных полосах

Крит. полоса	(размер WDFT = 256, $f_s = 32$ кГц)		
	Диапазон коэффициентов	Кол-во	Диапазон частот, Гц
1	1 – 4	4	22 – 86
2	5 – 9	5	108 – 195
3	10 – 13	4	217 – 283
4	14 – 18	5	305 – 395
5	19 – 23	5	417 – 509
6	24 – 28	5	533 – 628
7	29 – 33	5	653 – 752
8	34 – 39	6	778 – 910
9	40 – 45	6	937 – 1079
10	46 – 51	6	1109 – 1264
11	52 – 57	6	1297 – 1469
12	58 – 63	6	1506 – 1700
13	64 – 69	6	1741 – 1964
14	70 – 75	6	2012 – 2272
15	76 – 81	6	2339 – 2642
16	82 – 87	6	2711 – 3095
17	88 – 93	6	3182 – 3671
18	94 – 98	5	3782 – 4286
19	99 – 104	6	4428 – 5269
20	105 – 108	4	5469 – 6146
21	109 – 113	5	6402 – 7615
22	114 – 117	4	7973 – 9209
23	118 – 122	5	9680 – 11886
24	123 – 127	5	12517 – 15277
25	128 – 128	1	16000 – 16000

Основная задача кодера является разделение входного сигнала на частотные компоненты с помощью прямого WDFT. Далее, коэффициенты WDFT квантуются с учетом порога маскирования, рассчитанного в области коэффициентов WDFT (см. п.14.4.3). Распределение коэф-

коэффициентов WDFT в критических частотных полосах для частоты дискретизации 16кГц показано в табл. 14.2. Размер преобразования WDFT здесь равен 256. Анализ ошибки реконструкции сигнала звука показывает, что искажения, вносимые только процессом анализ – синтез (прямое – обратное WDFT), не воспринимаются на слух. На рис. 14.12 показан пример такого преобразования (аудио сигнал – кастаньеты).

Кодирование коэффициентов WDFT осуществляется с помощью кодовых книг Хаффмана [25]. Коды Хаффмана относятся к семейству кодов с переменной длиной кодового слова, т.е. каждый коэффициент кодируется последовательностью переменной длины бит. Формирование кодов Хаффмана базируется на знании закона распределения кодируемых данных.

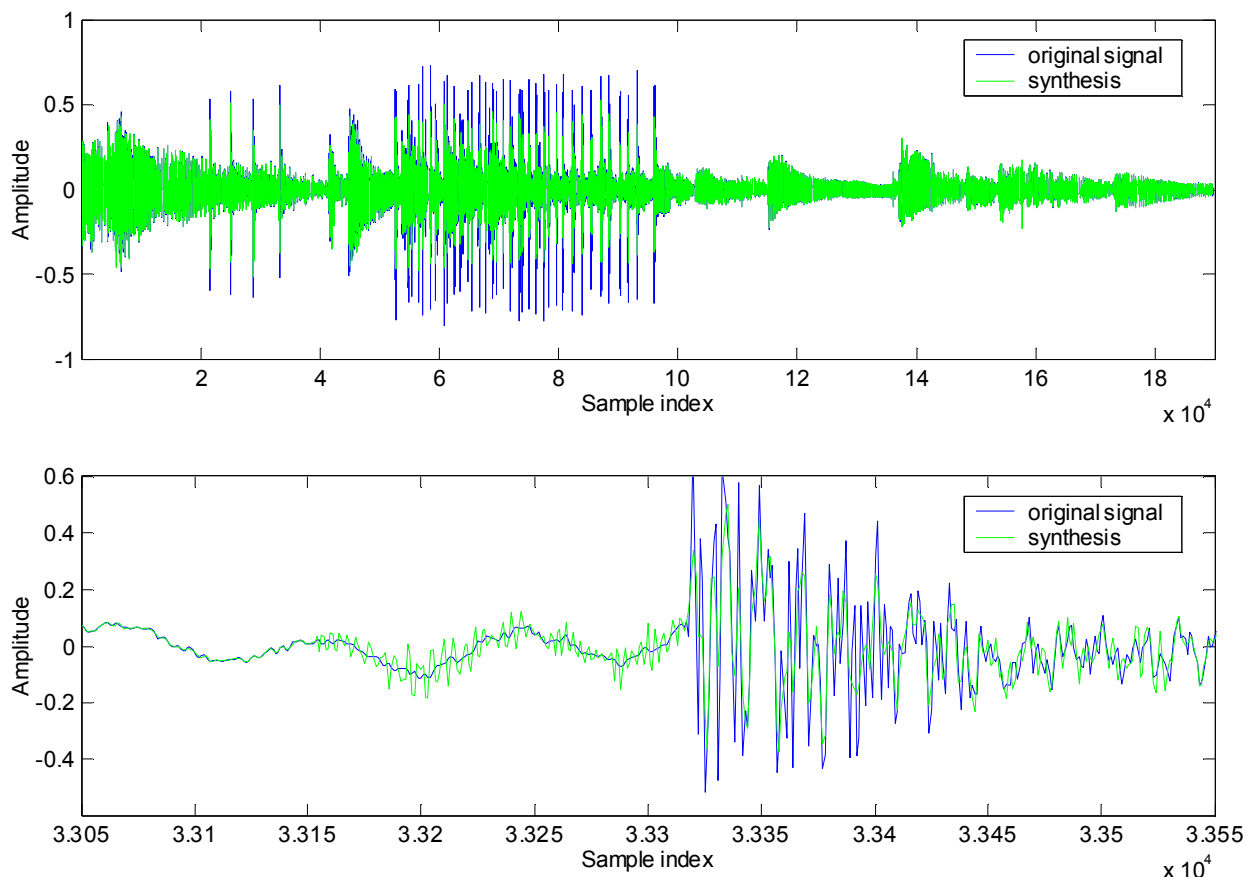


Рис. 14.12 – Пример обработки аудио сигнала: анализ – синтез (прямое – обратное WDFT)

Общее количество бит динамически распределяется в зависимости от энергии каждого спектрального компонента и его значимости. Пусть в какой-то момент времени спектральные компоненты сигнала обладают одинаковой энергией и занимают весь спектр, а также предполагается отсутствие модуля психоакустического анализа информации. Таким образом, все действия сконцентрированы на устранении статистической избыточности (далее просто избыточности). В данном случае, увеличение степени компрессии за счет перераспределения общего количества бит между всеми спектральными компонентами не осуществится в силу того, что для кодирования каждого компонента потребуется одно и то же количество бит. С другой стороны, если допустить, что спектр сигнала «окрашенный», например, основные спектральные компоненты сконцентрированы в области нижних частот, то произойдет перераспределение общего количества бит между всеми спектральными компонентами и значение степени ком-

прессии увеличится. Здесь сигнал содержит избыточность и соответственно в большей или в меньшей степени её можно устранить. Если положить, что x_k – k -я спектральная компонента сигнала, а $Q(x_k)$ – её R_k битный квантованный аналог, Q – операция квантования, тогда ошибка реконструкции k -ой компоненты равна $q_k = x_k - Q^{-1}(Q(x_k))$. Другими словами, q_k – внесенное искажение в сигнал в результате его кодирования. Оптимальное распределение бит по критическим частотным полосам определяется следующим образом [26], где N – количество критических частотных полос:

$$R_k = R + \frac{1}{2} \log_2(x_k^2) - \frac{1}{2} \log_2 \left(\prod_{k=0}^{N-1} x_k^2 \right)^{\frac{1}{N}}. \quad (14.34)$$

Выигрыш в количестве бит по сравнению с однополосным банком фильтров будет только в том случае, когда средне геометрическое значение спектральной плотности мощности сигнала будет много меньше её средне арифметического значения. Отношение средне геометрического значения спектральной плотности мощности сигнала к её средне арифметическому значению есть мера пологости спектра сигнала (Spectral Flatness Measure – *SFM*) [27].

$$SFM = \frac{(\prod_{k=0}^{N-1} x_k^2)^{\frac{1}{N}}}{\frac{1}{N} \sum_{k=0}^{N-1} x_k^2}. \quad (14.35)$$

Значения *SFM* варьируются в пределе от 0 до 1. Если *SFM* = 1, то подразумевается, что входной сигнал с пологим спектром и соответственно никакого увеличения компрессии нельзя получить. Пусть *SFM* = 1, тогда согласно (14.34) получается, что $R_k = R$. Таким образом, мерой избыточности в сигнале является мера пологости спектра *SFM* [27]: чем более пологий спектр сигнала, тем меньше избыточности в сигнале. Малое значение *SFM* подразумевает потенциально высокую степень компрессии сигнала, которую естественно можно оценить числом бит, необходимых для кодирования сигнала без артефактов.

В перцептуальном кодере сигналов стоит цель не только устранения информационной избыточности, но и изоляции перцептуальной избыточности акустической информации в сигнале. Это желание расположить в спектре сигнала внесенные искажения в реконструированный сигнал в результате квантования ниже порога маскирования, т.е. порога восприятия акустической информации слушателем. Соотношение сигнал шум *SNR* для квантования компонент каналов банка фильтров равно:

$$SNR = 10 \log \frac{x^2}{q^2}, \quad (14.36)$$

а соотношение сигнал к порогу маскирования *T SMR* определяется следующим образом

$$SMR = 10 \log \frac{x^2}{T^2}. \quad (14.37)$$

Далее для компонент сигнала k -го канала, значения которых больше порога маскирования T_k , хотелось бы максимизировать разность $SNR - SMR$, или, что эквивалентно, миними-

зировать разность $SMR - SNR$. Для соотношения $SMR - SNR$ с учетом дисперсии q^2 , дисперсия внесенных искажений кодированием, взвешенная маскирующим фактором, равна:

$$\frac{q^2}{T^2} = \frac{1}{N} \sum_{k=0}^{N-1} \left(\frac{x_k^2 / T_k^2}{3 \cdot 2^{2R_k}} \right), \quad (14.38)$$

где T_k – уровень порога маскирования в k -ом канале банка фильтров. Минимизация данной взвешенной ошибки (14.38), аналогично варианту минимизации дисперсии ошибки реконструкции q^2 , приводит к следующей формуле оптимального распределения бит по каналам банка фильтров:

$$R_k = R + \frac{1}{2} \log_2 \left(x_k^2 / T_k^2 \right) - \frac{1}{2} \log_2 \left(\prod_{k=0}^{N-1} x_k^2 / T_k^2 \right)^{\frac{1}{N}}. \quad (14.39)$$

Из (14.39) следует, что мера перцептуальной избыточности определяется как отношение:

$$PSFM = \frac{\left(\prod_{k=0}^{N-1} x_k^2 / T_k^2 \right)^{\frac{1}{N}}}{\frac{1}{N} \sum_{k=0}^{N-1} x_k^2 / T_k^2}. \quad (14.40)$$

Как видно, $PSFM$ зависит от распределения по частотному диапазону спектральной энергии взвешенной энергией порога маскирования.

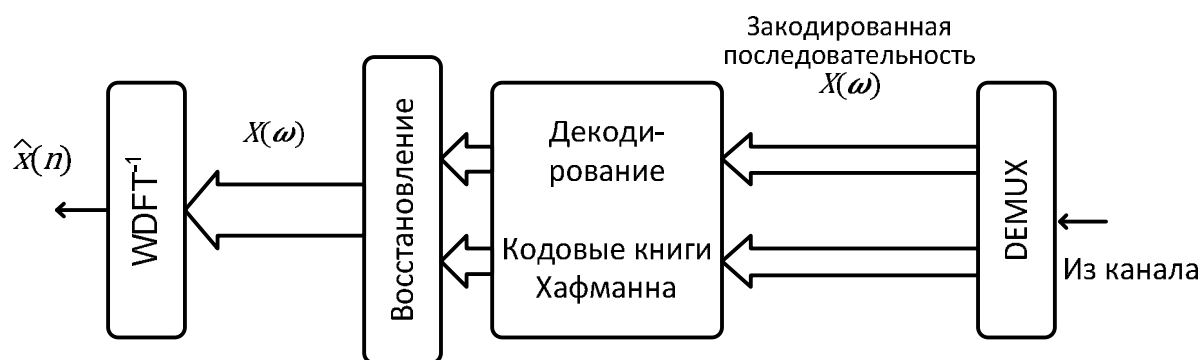


Рис. 14.13 – Структурная схема широкополосного перцептуального декодера речи и звука на основе WDFT

Итак, отношение среднеквадратического значения коэффициентов WDFT X_k в k -ой критической частотной полосе к соответствующему маскирующему порогу T_k , равномерно распределенному между коэффициентами WDFT X_k данной критической частотной полосы определяется следующим образом [6]:

$$SFM_k = \frac{|X_k|}{\sqrt{12 \cdot \frac{T_k}{K_k}}}, \quad (14.41)$$

где знаменатель $\sqrt{12 \cdot T_k/K_k}$ представляет собой максимальный шаг квантователя Δ_k коэффициентов WDFT в k -ой критической частотной полосе, а величина SFM_k задает минимальное количество уровней квантования (здесь K_k определяет число коэффициентов WDFT в k -ой критической частотной полосе согласно табл. 14.2).

Структура широкополосного перцептуального декодера (см. рис. 14.13) речи и звука на базе WDFT проще чем структура кодера и включает обратное преобразование WDFT, блоки восстановления и декодирования с минимальными потерями на основе кодовых книг Хаффмана.

14.6.2 Кодирование коэффициентов WDFT по Хаффману

В общем случае, коэффициенты WDFT в каждой критической частотной полосе имеют функцию распределения плотности вероятности Гаусса [6]:

$$f(x) = K \cdot e^{-\left(\frac{|x|}{\alpha}\right)^\beta} \quad (14.42)$$

Здесь $\beta = 2.0$ определяет функцию распределения плотности вероятности Гаусса. В соответствии с критерием согласия χ^2 функция распределения плотности вероятности Гаусса ближе аппроксимирует распределение коэффициентов WDFT в критической частотной полосе (табл. 14.3).

Табл. 14.3 - Параметры функции распределения плотности вероятности

Параметры	Номер группы l			
	1	4	5	6
α	0.001	0.08	0.13	0.18
β	2	2	2	2

Вероятность уровня квантования $qL_{k,j} = 2 \cdot \lfloor n \text{int}(SMR_{k,j}) \rfloor + 1$ в k -ой критической частотной полосе определяется следующим образом:

$$p(qL_{k,j}) = \frac{1}{2} e^{-\left(\frac{qL_{k,j} \cdot \delta_k}{\alpha_l}\right)^{\beta_l}} \sinh\left(\frac{\delta_k}{\alpha_l}\right), qL_{k,j} \neq 0, \quad (14.43)$$

$$p(0) = 1 - e^{-\left(\frac{\delta_k}{2 \cdot \alpha_l}\right)^{\beta_l}},$$

где α_l и β_l - параметры функции распределения плотности вероятности из табл. 14.3; l – номер группы соответствующей числу коэффициентов k -полосы; δ_k - шаг квантования коэффициентов WDFT k -ой критической частотной полосы.

Далее, операция кодирования уровней квантования $qL_{k,j}$ каждого коэффициента $X_{k,j}$ на основе адаптивных кодовых книг Хаффмана выполняется так: каждому уровню квантования ставится в соответствие из кодовой книги бинарный вектор $B_{k,j}$ длины w_k

$$B_k = (b_{k,1}, b_{k,2}, b_{k,3}, \dots, b_{k,w_k}), b_{k,j} \in \{0,1\}, j = 1, \dots, w_k. \quad (14.44)$$

Причём, очевидно, вероятность данного вектора равна вероятности уровня квантования $p(B_{k,j}) = p(qL_{k,j})$. Минимально достижимое кодовое слово R_k в k -й полосе по Хаффману вычисляется как

$$R_k = \sum_j p(B_{k,j}) \cdot w_j. \quad (14.45)$$

При этом, среднее значение длины кодового слова должно быть не меньше значения энтропии входной последовательности

$$R_k \geq H(p(qL_{k,j})) = \sum_k p(qL_{k,j}) \cdot \log_2 \left(\frac{1}{p(qL_{k,j})} \right). \quad (14.46)$$

Табл. 14.4 – Кодовая книга Хаффмана для группы $l = 6$

qL_k	$p(qL_k)$	Кодовое слово B_k	Длина слова w_k	$p(qL_k)w_k$	qL_k	$p(qL_k)$	Кодовое слово B_k	Длина слова w_k	$p(qL_k)w_k$
0	0.1071	011	3	0.3214					
1	0.1019	001	3	0.3056	-1	0.1019	010	3	0.3056
2	0.0898	1110	4	0.3594	-2	0.0898	1111	4	0.3594
3	0.0744	1100	4	0.2974	-3	0.0744	1011	4	0.2974
4	0.0580	1000	4	0.2322	-4	0.0580	1001	4	0.2322
5	0.0429	0000	4	0.1715	-5	0.0429	11011	5	0.2144
6	0.0300	10100	5	0.1502	-6	0.0300	00011	5	0.1502
7	0.0200	110101	6	0.1200	-7	0.0200	110100	6	0.1200
8	0.0127	000101	6	0.0761	-8	0.0127	000100	6	0.0761
9	0.0077	1010101	7	0.0536	-9	0.0077	1010100	7	0.0536
10	0.0044	10101110	8	0.0353	-10	0.0044	10101101	8	0.0353
11	0.0024	101011110	9	0.0219	-11	0.0024	101011001	9	0.0219
12	0.0013	1010111111	10	0.0128	-12	0.0013	1010111110	10	0.0128
13	0.0006	1010110000	10	0.0064	-13	0.0006	10101100011	11	0.0071
14	0.0003	101011000100	12	0.0037	-14	0.0003	101011000101	12	0.0037
								$R = 4.0570$	бит/отсчет
								$H(p(qL_k)) = 4.0189$	бит/отсчет
								$cr = 0.9484$	%

Близость среднего значения длины кодового слова R_k к значению энтропии $H(p(qL_{k,j}))$ в (14.46), говорит о приближении к оптимальным кодам представления входных

данных. Кодирование будет оптимальным только тогда, когда значение средней длины кодового слова будет равно значению энтропии $R_k = H(p(qL_{k,j}))$. Однако такое представление на практике не всегда возможно, поэтому избыточность кодового представления коэффициентов WDFT по Хаффману оценивается согласно формуле

$$cr_k = \frac{R_k - H(p(qL_{k,j}))}{H(p(qL_{k,j}))} \cdot 100\%. \quad (14.47)$$

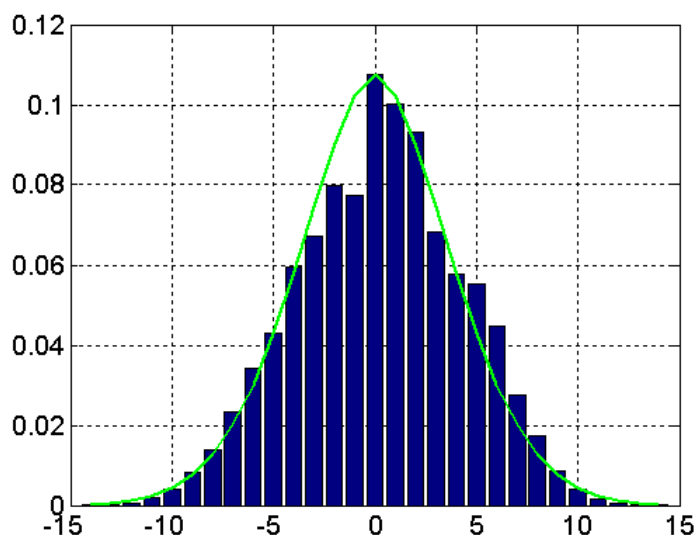


Рис. 14.14 – Гистограмма распределения вероятности коэффициентов WDFT и аппроксимирующая функция распределения вероятности Лапласа (непрерывная кривая) с параметрами $\beta = 2$, $\alpha = 0.18$.

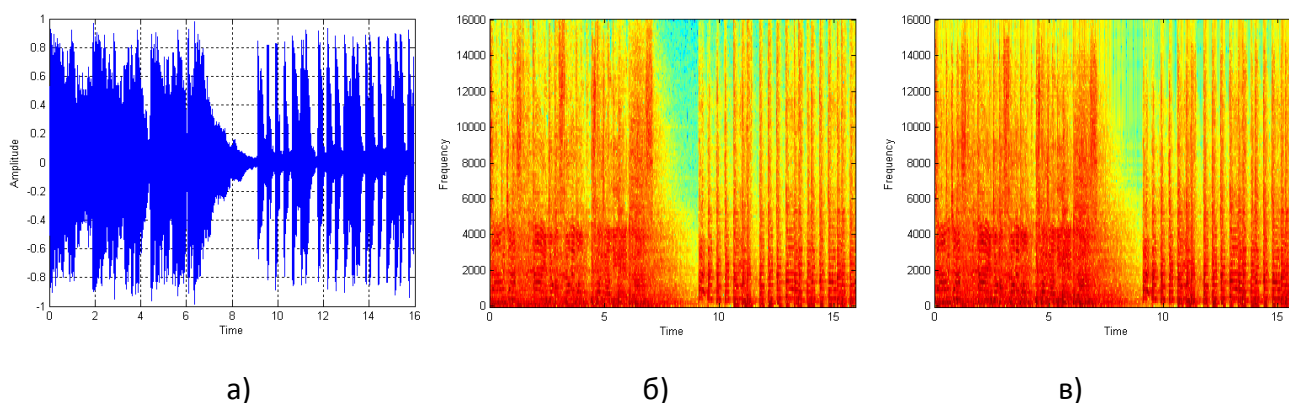


Рис. 14.15 – Пример обработки музыкального сигнала: а) исходный сигнал; б) спектрограмма исходного сигнала; в) спектрограмма реконструированного сигнала.

14.6.3 Экспериментальные результаты

Описанный выше широкополосный перцептуальный кодер речи и звука для входных сигналов с частотами дискретизации 16 и 32 кГц характеризуется скоростью передачи не более чем 24 кБит/с и 46 кБит/с соответственно. Применение адаптивного кодирования на основе кодовых книг Хаффманна с Гауссовской моделью распределения, позволяет обеспечить 1.45 бита на отсчет входного сигнала для обоих типов сигналов.

На рис. 14.15 показан исходный сигнал с частотой дискретизации 32 кГц, его спектрограмма и реконструированный сигнал на выходе декодера. Качество восстановленного сигнала подтверждается субъективными экспертными оценками, говорящими о прозрачном или почти прозрачном кодировании данных.

14.7 Список использованных источников

1. Bagchi S., Mitra S.K., The Nonuniform Discrete Fourier Transform and Its Applications in Filter Design, IEEE Trans. Circuits Systems II, 1996, Vol. 43, No.6, pp. 422–433.
2. Nam Cho Ik., Mitra S.K., Warped discrete cosine transform and its application in image compression, IEEE Trans. Circ. Systems Video Technol, 2000, Vol. 10, No. 8, pp. 1364-1373.
3. Makur A., Mitra S.K., Warped Discrete-Fourier Transform: Theory and Applications, IEEE Trans. Circuits Systems I, 2001, Vol. 48, No. 9, pp. 1086–1093.
4. Borowicz A., Parfieniuk M., Petrovsky A.A., An application of the warped discrete Fourier transform in the perceptual speech enhancement, Speech Communication. ELSEVIER, 2006, Vol. 48, pp. 1024-1036.
5. Parfieniuk M., Petrovsky A., Warped DFT as the basis for psychoacoustical model, Acoustics, speech and signal processing: Proc. ICASSP, Vol. 4, Montreal, Canada, 2004. pp. 185–188.
6. Petrovsky A.A. and et al., Warped discrete Fourier transform in perceptual speech and audio processing, Proc. "X Symposium New Trends In Audio and Video", Wroclaw, Poland, 2004, pp. 143-152.
7. Jayant N.S., Chen E.Y., Audio compression: technology and applications, AT&T technical journal, Vol. 74, No 2, 1995, pp. 23-34.
8. Boll S.F., Suppresion of acoustic noise in speech using spectral subtraction, IEEE Transaction Acoustic, Speech, Signal Processing, 1979, Vol. 27, No. 2, pp. 113-120.
9. Ephraim Y., Malah D., Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, IEEE Transaction on Acoustic, Speech, and Signal Processing, 1984, Vol. ASSP-32, No. 6, pp. 1109–1121.
10. Virag N., Single channel speech enhancement based on masking properties of the human auditory system, IEEE Trans. on Speech and Audio Processing, 1999, Vol. 7, No. 2, pp. 126–137.
11. Petrovsky A.A., Anoshenko A.E., Combined system for echo cancelation and noise reduction in frequency domain with psychoacoustic motivation, The 2nd International conference and exhibition on Digital Signal Processing and its Applications, DSPA-99, Moscow, Russia, 1999, pp. 166-169.
12. Gustafsson S. and et al., A novel psychoacoustically motivated audio enhancement algorithm preserving background noise characteristic, IEEE Int. Conf. on Acoustic, Speech, and Signal Processing ICASSP'98, Seattle, USA, 1998, Vol. 1, pp. 397–400.

-
13. Johnston J.D., Transform coding of audio signals using perceptual noise criteria, *IEEE Transactions on Selected Areas Communication*, February, 1988, Vol. 6, pp. 314–323.
 14. Petrovsky A.A., Parfieniuk M., Borowicz A., Warped DFT based perceptual noise reduction system, AES, Convention Paper #6035, presented at the 116th Convention, 2004 May 8-11 Berlin, Germany, 16 pages.
 15. Петровский А.А., Борович А., Парфенюк М., Обработка речи на основе дискретного преобразования Фурье с неравномерным частотным разрешением, *Речевые технологии*, № 3, Москва, 2008, с. 3-15.
 16. Borowicz A., Petrovsky A.A., The comparative study of voice activity detectors, *VII Int. Conf. of Modern Telecommunication Systems, Naroch*, 2002, pp. 148-152.
 17. Martin R., Spectral subtraction based on minimum statistics, *VII European Signal Processing Conf. EUSIPCO'94*, Edinburg, 1994. pp. 1182-1185.
 18. Martin R., Noise power spectral density estimation based on optimal smoothing and minimum statistics, *IEEE Transaction on Speech and Audio Processing*, Vol. 9, No. 5, 2001, pp. 504-512.
 19. Cohen I., Berdugo B., Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement, *IEEE Signal Processing Letters*, Vol. 9, No. 1, January 2002, pp. 12-15.
 20. Petrovsky A.A., Bielawski K., Auditory model based enhancement system for hands-free devices, *XI European Signal Proc. Conf. (EUSIPCO'2002)*, Vol. 1, Toulouse, France, 2002, pp. 487-490.
 21. Piotrowski A., Parfieniuk M., *Cyfrowe banki filtrow: analiza, synteza I implementacja dla systemow multimedialnych*, Wydawnictwo Politechniki Bialostockiej, Bialystok, 2006. – 389s.
 22. Livshitz M., Petrovsky A., An overcomplete WDFT-based perceptually constrained variable bit rate wideband speech coder with embedded noise reduction system, *In Proc. XI Int. Conf. Speech and Computer (SPECOM'2006)*, St. Petersburg, Russia, 25-29 June, 2006, pp. 343-348.
 23. Yang W., Benbouchta M., Yantorno R., Performance of the modified bark spectral distortion as an objective speech quality measure, *Proc. of IEEE ICASSP, USA, Seattle*, May 12–15, 1998, pp. 541–544.
 24. Petrovsky Al. and et al., Auditory modeling via frequency warped transforms, *Proc. of the 124th AES Convention*, 2008. Preprint 8976. Amsterdam, Netherlands, 15 pages.
 25. Huffman D., A method for the construction of minimum redundancy codes, *The Proc. of IRE.*, Sept. 1952, Vol. 40, pp. 1098-1101.
 26. Петровский Ал., Белявский К., Петровский А.А., Перцептуальное кодирование аудио и речевых сигналов, *Доклады БГУИР № 1 (5) 23 февраля 2004*. с. 73-91.
 27. Spanias A., Painter T., Atti V., *Audio signal processing and coding*, Hoboken, NJ: John Wiley & Sons, Inc., 2007, 464 p.

15 ПРОЕКТИРОВАНИЕ БАНКА ФИЛЬТРОВ ДЛЯ СЛУХОВЫХ АППАРАТОВ С ИСПОЛЬЗОВАНИЕМ ЧАСТОТНОГО РАСТЯЖЕНИЯ И ОБЪЕДИНЕНИЯ СУБПОЛОС

Парфенюк М., Вашкевич М.И., Петровский А.А.

АННОТАЦИЯ

В главе рассматриваются вопросы проектирования неравнополосного косинусно-модулированного банка фильтров для слуховых аппаратов с использованием частотного растяжения и объединения субполос. Представлена полифазная структура для реализации рассматриваемого банка фильтров. Уделено внимание проектированию фильтра-прототипа, а также выбору коэффициентов децимации субполосных сигналов. Показано несколько подходов к решению задачи выравнивания групповой задержки в банке фильтров. Проработан вопрос объединения субполос в банке фильтров для увеличения точности аппроксимации критических частотных полос. Для оценки точности аппроксимации вводятся два критерия, основанные на среднеквадратичной и максимальной ошибках отклонения. Кроме того рассмотрены банки фильтров, основанные на использовании фазовых звеньев без операций умножения.

Ключевые слова – банк фильтров, объединение субполос, фильтр-прототип, частотное растяжение.

СОДЕРЖАНИЕ

15.1 Введение.....	331
15.2 Косинусно-модулированный банк фильтров с деформированной частотной осью.....	333
15.2.1 Механизм неравнополосного частотного разбиения	333
15.2.2 Полифазная структура банка фильтров	334
15.2.3 Выбор коэффициентов децимации ...	337
15.2.4 Выравнивание групповой задержки..	337
15.2.5 Расчет фильтра-прототипа	341
15.3 Объединение субполос.....	342
15.3.1 Обоснование	342
15.3.2 Точность частотного разбиения.....	343
15.4 Неравнополосный косинусно-модулированный банк фильтров с частотной деформацией и объединением субполос на основе фазовых звеньев без умножителей	345
15.5 Список использованных источников.....	350

15.1 Введение

Слуховые аппараты помогают большому количеству людей, потерявших по тем или иным причинам способность нормально слышать, вернуться к трудовой и общественной дея-

тельности [1]. Первые слуховые аппараты (СА) представляли собой аналоговые системы, которые использовали примитивные концепции для компенсации потери слуха. В таких системах было сильно ограничено число степеней свободы, которое необходимо, чтобы справиться с большим разнообразием нарушений слуха. Революционным прорывом был переход от аналоговых СА к цифровым. Последние обладают высокой гибкостью, которая характерна для цифровых систем. Переход от аналоговых к цифровым СА явился предпосылкой к разработке большого количества алгоритмов, которые используются в СА для точной настройки под конкретного пользователя [2].

На рис. 15.1 изображена схема современного слухового аппарата. Акустический сигнал поступает в устройство с нескольких микрофонов, из которых формируется моносигнал. В дальнейшем в СА ведется субполосная обработка полученного моносигнала. Важнейшими шагами субполосной обработки являются редактирование шума и усиление отдельных субполос для корректировки частотной характеристики слуха пациента. Технически сложной является задача компенсации эхо-сигнала. Почти все упомянутые компоненты СА могут быть различным образом настроены в зависимости от ситуации. Для подобной настройки необходима система классификации, в которой анализируется акустический сигнал. Анализ заключается в извлечении признаков, по которым классифицируется в какой акустической обстановке находится пользователь. Это влечет за собой изменение параметров, которые регулируют блок редактирования шумов и модуль усиления субполосных сигналов.

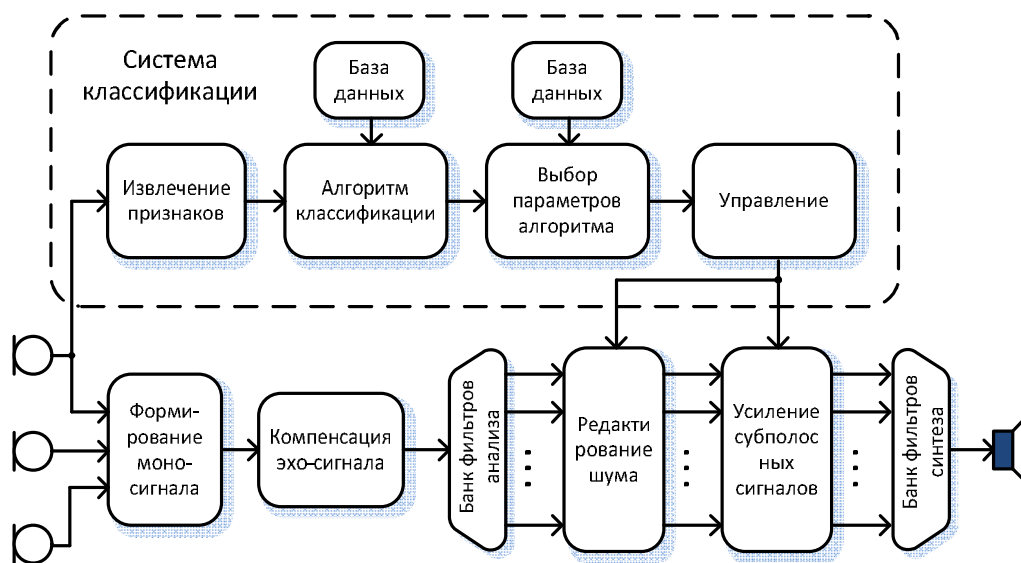


Рис. 15.1 – Стадии обработки сигнала в слуховом аппарате

Качество работы слухового аппарата во многом определяется способом декомпозиции сигнала на субполосные составляющие. Поэтому в последнее время в литературе появляется множество публикаций, которые посвящены методам проектирования банков фильтров для слуховых аппаратов [3–5]. В данной главе будет рассмотрен банк фильтров для слуховых аппаратов, использующий частотное растяжение и объединение субполос [6].

В отличие от многих авторов [7–8], которые фокусируются на исследовании неравнополосных ДПФ-модулированных банков фильтров, мы будем рассматривать банки фильтров основанные на косинусной модуляции [9]. Их преимущество в том, что для действительных

входных сигналов каналные сигналы также являются действительными, что значительно облегчает их последующую обработку и техническую реализацию банка фильтров.

15.2 Косинусно-модулированный банк фильтров с деформированной частотной осью

15.2.1 Механизм неравнополосного частотного разбиения

Неравнополосный косинусно-модулированный банк фильтров на основе деформации оси частот выводится из своего равнополосного аналога [9]. При этом для последнего импульсные характеристики фильтров анализа и синтеза имеют следующий вид:

$$\begin{aligned} h_k(n) &= 2p(n) \cos \left[(2k+1) \frac{\pi}{2M} \left(n - \frac{N-1}{2} \right) + (-1)^k \frac{\pi}{4} \right], \\ f_k(n) &= 2p(n) \cos \left[(2k+1) \frac{\pi}{2M} \left(n - \frac{N-1}{2} \right) - (-1)^k \frac{\pi}{4} \right], \\ &0 \leq k < M, \end{aligned} \quad (15.1)$$

где $p(n)$ — это коэффициенты фильтра-прототипа порядка N с частотой среза $\omega_c = \pi/2M$, а M количество каналов банка фильтров. На порядок фильтра-прототипа накладывается ограничение $N = 2mM$, где m — произвольное положительное число. Далее для получения неравнополосного банка фильтров применяется фазовое преобразование (англ. allpass transform), которое заключается в замене в системе всех элементов задержки z^{-1} на фазовые звенья $A(z)$. Получаемые при этом передаточные функции фильтров анализа и синтеза M -канального косинусно-модулированного банка фильтров, основанного на деформации оси частот описываются следующими выражениями [9,10]:

$$\begin{aligned} H_k(z) &= a_k b_k U_k(A^{-1}(z)) + \bar{a}_k \bar{b}_k V_k(A^{-1}(z)), \\ F_k(z) &= \bar{a}_k b_k U_k(A^{-1}(z)) + a_k \bar{b}_k V_k(A^{-1}(z)), \end{aligned} \quad (15.2)$$

где верхние черточки означают комплексное сопряжение,

$$a_k = e^{j(-1)^k \frac{\pi}{4}}, \quad (15.3)$$

$$b_k = W_{2M}^{-\left(k+\frac{1}{2}\right)\left(\frac{L-1}{2}\right)}, \quad (15.4)$$

$$U_k(z) = P \left(z W_{2M}^{k+\frac{1}{2}} \right), \quad (15.5)$$

$$V_k = P \left(z W_{2M}^{-\left(k+\frac{1}{2}\right)} \right), \quad (15.6)$$

$W_{2M} = e^{\frac{j\pi}{M}}$, $P(z)$ — z -преобразование фильтра-прототипа. Для расчета фильтра-прототипа могут использоваться методы, разработанные для равнополосных банков фильтров, см. [11–12] и литературу к этим статьям.

Простая замена $z^{-1} \rightarrow A(z)$ имеет примечательный эффект: частотная характеристика равнополосного банка фильтров деформируется в соответствии с фазовой характеристикой $\varphi(\omega)$ как показано на рис. 15.2. Простейшая деформация оси частот основывается на применении передаточной функции фазового звена первого порядка

$$A(z) = \frac{\alpha + z^{-1}}{1 + \alpha z^{-1}}, \quad (15.7)$$

фазовая характеристика $\varphi(\omega)$ которого зависит только от действительного коэффициента $|\alpha| \leq 1$. Чем меньше абсолютное значение коэффициента, тем слабее деформация оси частот. Для $\alpha = 0$ фактически не происходит деформации оси частот, т.к. $A(z) = z^{-1}$, т.е. это единичная задержка. За исключением данного вырожденного случая, для реализации такого фильтра требуется одно умножение и два элемента задержки, либо два умножения и один элемент задержки.

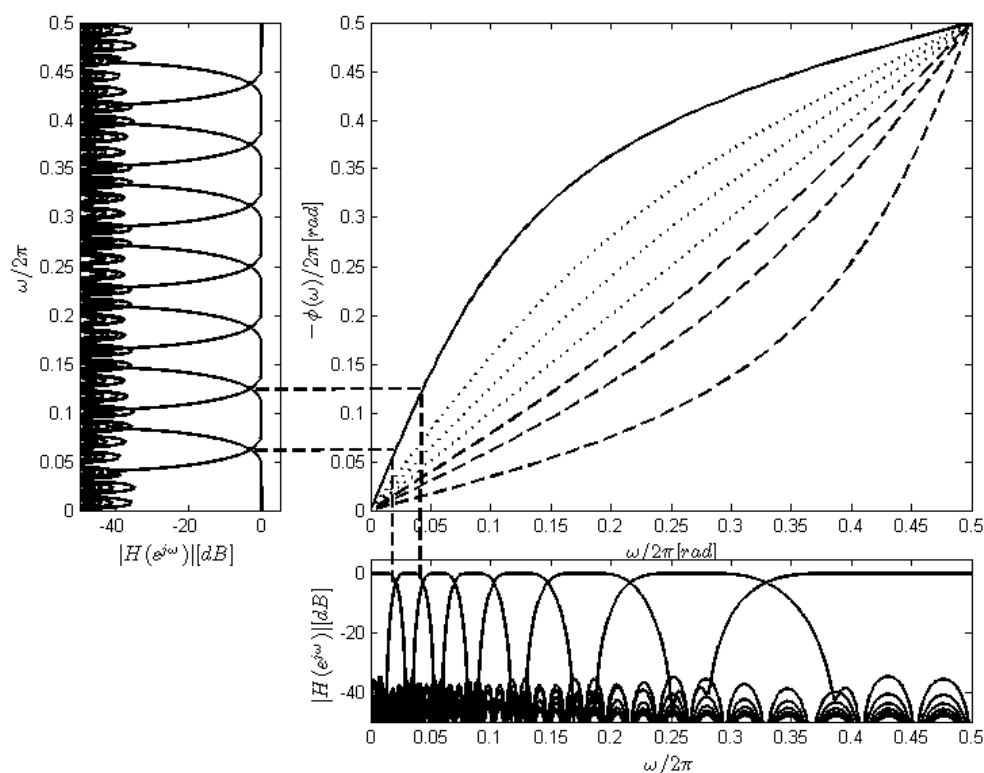


Рис. 15.2 – Фазовое преобразование косинусно-модулированного банка фильтров

15.2.2 Полифазная структура банка фильтров

В работе [13] показано, что косинусно-модулированный банк фильтров можно эффективно реализовать, используя полифазное представление фильтра-прототипа:

$$P(z) = \sum_{n=0}^{N-1} p(n) z^{-n} = \sum_{q=0}^{2M-1} \sum_{r=0}^{m-1} p(q + 2rM) z^{-(q+2rM)} = \sum_{q=0}^{2M-1} z^{-q} G(z^{2M}), \quad (15.8)$$

где $G(z)$ это полифазные компоненты I типа.

Введем обозначение

$$c_{k,l} = 2 \cos \left(\frac{(2k+1)\pi}{2M} \left(l - \frac{N-1}{2} \right) + (-1)^k \frac{\pi}{4} \right), \quad (15.9)$$

из которого, используя периодичность функции косинус, получаем тождество:

$$c_{k,(l+2rM)} = (-1)^r c_{k,l}. \quad (15.10)$$

Объединяя полифазное представление фильтра-прототипа (15.8) с выражениями (15.9)–(15.10), получаем формулу для эффективной реализации косинусно-модулированного банка фильтров:

$$H_k(z) = \sum_{n=0}^{N-1} h_k(n) z^{-n} = \sum_{n=0}^{N-1} p(n) c_{k,n} = \sum_{q=0}^{2M-1} \sum_{r=0}^{m-1} p(q + 2rM) c_{k,(q+2rM)} z^{-(q+2rM)}. \quad (15.11)$$

После чего, применяя упрощение (15.9) к (15.11), получим:

$$H_k(z) = \sum_{n=0}^{N-1} z^{-q} c_{k,q} \sum_{r=0}^m (-1)^r p(q + 2rM) z^{-2rM} = \sum_{q=0}^{2M-1} c_{k,q} G_q(-z^{2M}). \quad (15.12)$$

Данное выражение можно переписать в матричной форме:

$$\begin{bmatrix} H_0(z) \\ H_1(z) \\ \vdots \\ H_{M-1}(z) \end{bmatrix} = \mathbf{C}_A \times \begin{bmatrix} G_0(-z^{2M}) \\ z^{-1} G_1(-z^{2M}) \\ \vdots \\ z^{-(2M-1)} G_{2M-1}(-z^{2M}) \end{bmatrix}, \quad (15.13)$$

здесь $\mathbf{C}_A = [c_{k,l}]_{0 \leq k < M, 0 \leq l < 2M}$ это матрица косинусной модуляции, которая, как было показано в [13], сводится к вычислению M -точечного дискретного косинусного преобразования типа IV (ДКП-IV).

Аналогично, представляя фильтр-прототип в виде суммы полифазных компонент II типа

$$H(z) = \sum_{q=0}^{2M-1} z^{-(2M-1-q)} R_q(z^{2M}),$$

выводятся формулы для эффективной реализации банка фильтров синтеза:

$$\begin{bmatrix} F_0(z) \\ F_1(z) \\ \vdots \\ F_{M-1}(z) \end{bmatrix}^T = \begin{bmatrix} z^{-(2M-1)} R_0(-z^{2M}) \\ z^{-(2M-2)} R_1(-z^{2M}) \\ \vdots \\ R_{2M-1}(-z^{2M}) \end{bmatrix} \times \mathbf{C}_S, \quad (15.14)$$

где $\mathbf{C}_S = \mathbf{C}_A^T$.

На основании выражений (15.13) и (15.14) строится полифазная структура банка анализа и синтеза соответственно. Далее, чтобы получить неравнополосный косинусно-

модулированный банк фильтров необходимо в полифазной структуре заменить все элементы задержки на фазовые звенья т.е. $z^{-1} \rightarrow A(z)$. Получаемая после этого структура косинусно-модулированного банка фильтров с деформированной осью частот показана на рис. 15.3. Нужно отметить, что на данном рисунке через $g_i, i = 0 \dots N - 1$ обозначены коэффициенты полифазных компонент фильтра-прототипа, которые находятся исходя из формулы (15.12).

Отличие данной полифазной структуры от той, что применяется для равнополосных банков фильтров в том, что все элементы задержки заменены фазовыми звеньями, частота дискретизации канальных сигналов изменяется за пределами банка фильтров, а после реконструкции сигнала следует фильтр $C(z)$, выравнивающий групповую задержку. Данные особенности усложняют техническую реализацию банка фильтров и являются негативными последствиями, которые сопровождают полезный эффект деформации оси частот.

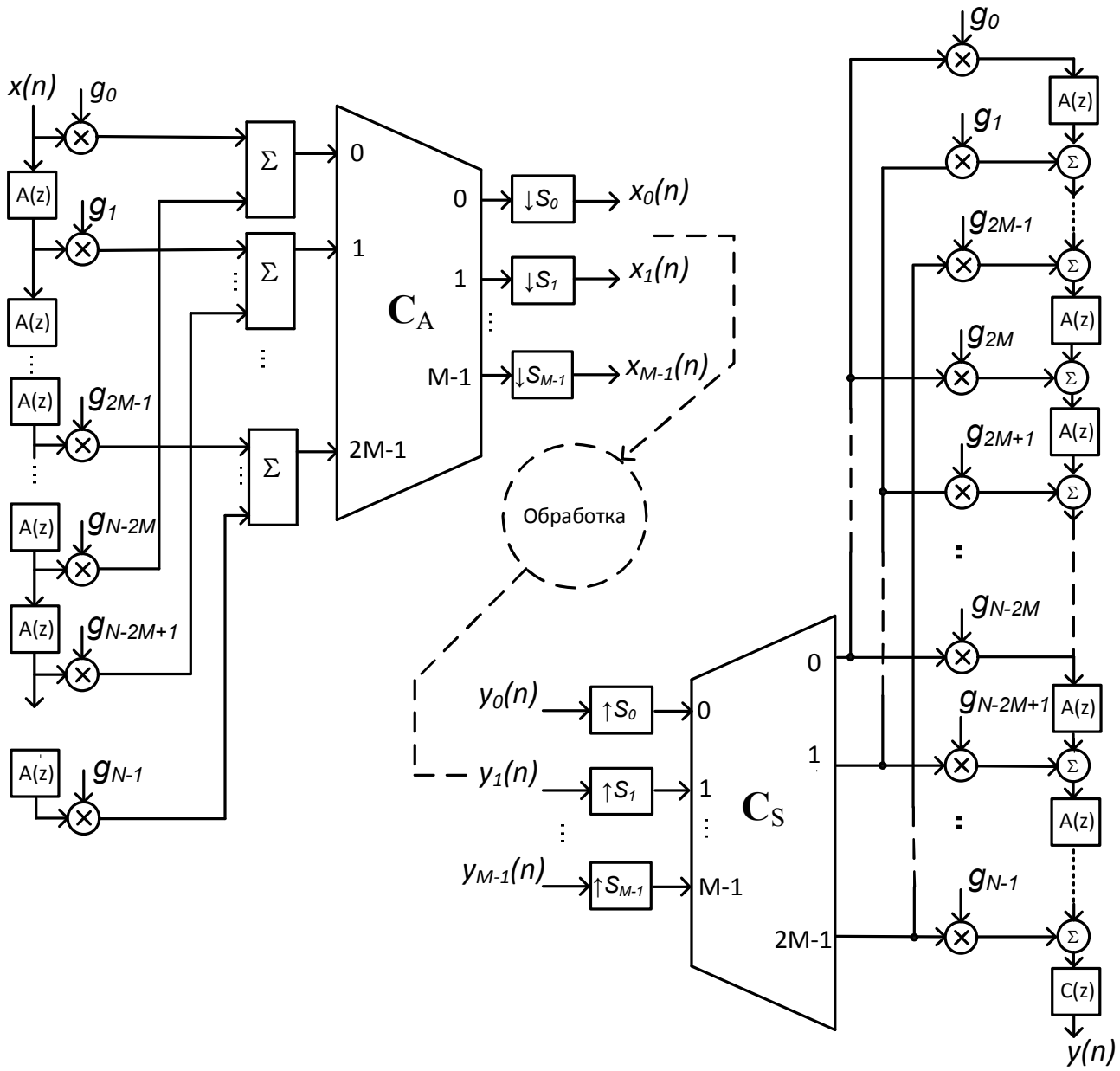


Рис. 15.3 – Структура косинусно-модулированного банка фильтров с деформированной осью частот

15.2.3 Выбор коэффициентов децимации

В системах на основе деформации оси частот предельная децимация без внесения искажений невозможна. Тем не менее, полная реконструкция хорошо аппроксимируется путем выбора избирательного фильтра-прототипа для ослабления компонент наложения спектров. Кроме этого для предотвращения появления алайзинговых эффектов требуется наложить ограничения на коэффициенты децимации канальных сигналов в соответствии с их шириной полосы и расположением на частотной оси.

В работе [14] были выведены следующие неравенства:

$$\left\lfloor \frac{n_k}{2f_{U_k}} f_s \right\rfloor \geq S_k \geq \left\lfloor \frac{n_k - 1}{2f_{L_k}} f_s \right\rfloor, \quad 1 \leq n_k \leq \left\lfloor \frac{f_{U_k}}{f_{U_k} - f_{L_k}} \right\rfloor, \quad (15.15)$$

в которых полагается что сигнал k -ого канала является действительным, занимает частотный диапазон $[-f_{U_k}, -f_{L_k}] \cup [f_{L_k}, f_{U_k}]$ и имеет частоту дискретизации f_s . Выражения определяют диапазон значений коэффициента децимации S_k , для которых не возникает значительного, необратимого наложения спектров. Разумеется, что наибольший интерес представляет максимальное допустимое значение S_k . Также следует отметить, что нижняя и верхняя частоты f_{L_k} и f_{U_k} , соответственно, должны включать (хотя бы частично) переходную полосу фильтров анализа, образуемую после деформации частотной оси.

15.2.4 Выравнивание групповой задержки

Наиболее важной проблемой использования банков фильтров в слуховых аппаратах, работающих в реальном времени, является то, что субполосная система анализ/синтез вносит задержку. Для того, чтобы пользователя не беспокоил эхо-сигнал его собственной речи, или не было несоответствия между движением губ говорящего и словами, которые слышит пользователь, временной сдвиг между входным сигналом и обработанным должен удерживаться в определенных рамках.

Упомянутые временные границы зависят от конкретного человека и поэтому затруднительно дать их определенное значение. Имеющиеся работы различаются в этом отношении [3], [15]. Общее заключение представляется таким, что задержка свыше 10 мс объективно вызывает эхо, в то время как задержка выше 20–30 мс делает не пригодной систему для использования в двухстороннем общении, потому что теряется согласованность восприятия с движением губ.

Растяжение частотной оси приводит к тому, что групповая задержка становится нелинейной функцией, поскольку банк фильтров на основе фазового преобразования всегда является БИХ системой. Групповая задержка зависит от силы деформации частотной оси и является монотонной функцией, с направлением определяемым знаком коэффициента фазового преобразования. Для $\alpha < 0$ нижние частоты задерживаются на число отсчетов большее порядка фильтра, в то время как высокие частоты задерживаются меньше. Для $\alpha > 0$, ситуация изменяется в противоположную сторону. Заметим также, что отклик системы анализ/синтез с деформированной осью частот эквивалентен отклику цепочки фазовых звеньев из N секций [16].

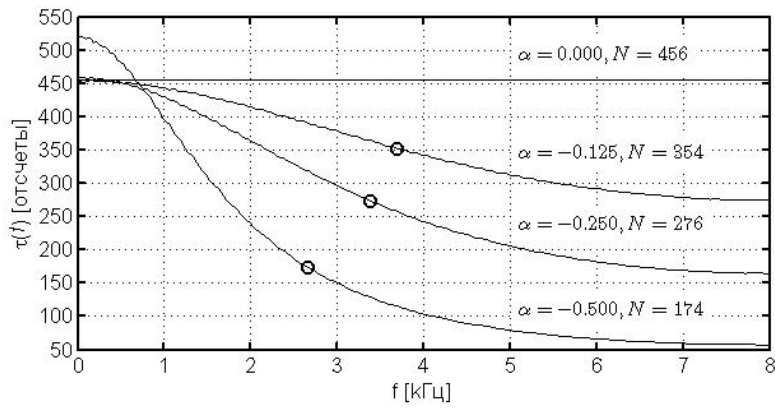


Рис. 15.4 – График групповой задержки для различных банков фильтров

На рис. 15.4 показаны групповые задержки банков фильтров с различными фильтрами прототипами (порядка N) и различными коэффициентами фазового преобразования (α). Кружками обозначены точки, в которых групповая задержка такая же, как у соответствующих равнополосных банков фильтров (с таким же фильтром-прототипом). Легко заметить, что отличие между этими значениями является существенным, поэтому остро стоит вопрос выравнивания групповой задержки в системах на основе деформации оси частот.

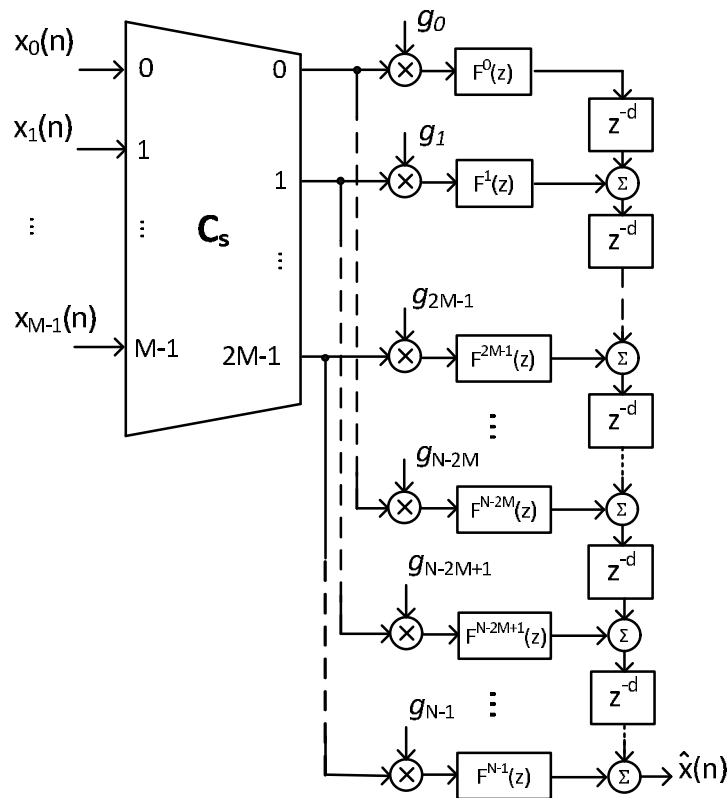


Рис. 15.5 – Банк фильтров синтеза

Как известно, групповая задержка линейной системы $\tau(\omega)$ равна производной от ФЧХ системы с обратным знаком:

$$\tau(\omega) = -\frac{d\varphi(\omega)}{\omega}, \quad (15.16)$$

из этого следует, что изменение групповой задержки в зависимости от частоты (рис. 15.4) это следствие нелинейности ФЧХ системы анализ/синтез неравнополосного косинусно-модулированного банка фильтров. Нелинейная ФЧХ системы недопустима, поскольку банк фильтров предполагается использовать для обработки звуковых сигналов. Поэтому в данном разделе рассмотрено несколько подходов к решению этой проблемы.

Первый рассматриваемый метод был предложен в [7]. В соответствии с данным методом банк фильтров синтеза строится, как показано на рис. 15.5.

На данной схеме $F(z)$ определяется следующим образом:

$$F(z) = -\alpha z^{-d} + (1 - \alpha^2)z^{-(d-2)} + \dots + \alpha^{d-2}(1 - \alpha^2), \quad (15.17)$$

где $d \in \mathbb{N}$. Данный подход позволяет проектировать неравнополосные банки фильтров с незначительными искажениями. Тем не менее, практическое применение данного подхода сдерживает его высокая вычислительная сложность. Очевидно, что функция $F(z)$ вычисляется $\frac{N}{2}(N - 1)$ раз для каждого отсчета. Более того, порядок $F(z)$ напрямую зависит от параметра d (формула (15.17)), который может быть весьма большим для хорошей компенсации фазовых искажений.

Как было показано в [16], количество вычислений в схеме на рис. 15.5 можно уменьшить, если её преобразовать следующим образом (рис. 15.6).

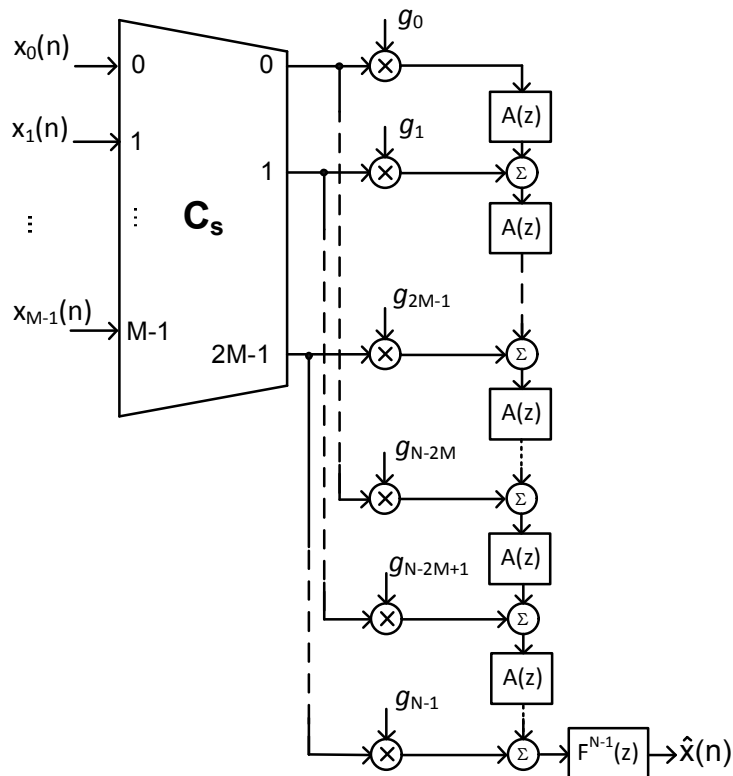


Рис. 15.6 – Улучшенный вариант схемы на рис. 15.5

Табл. 15.1 – Вычислительная сложность и затраты памяти

Структура	Операций умножения	Кол-во ячеек памяти
Рис. 15.5	$2N + N(N - 1)d$	$N(N - 1) + 2(N - 1)d$
рис. 15.6	$2N + 4(N - 1) + 2(N - 1)d$	$2(N - 1) + 2(N - 1)d$

Передаточные функции схем на рис. 15.6 и рис. 15.5 равны, однако количество вычислений во второй схеме значительно меньше. В табл. 15.1 приведены сведения по количеству необходимых операций умножения и затратам памяти для двух приведенных схем построения банка фильтров синтеза.

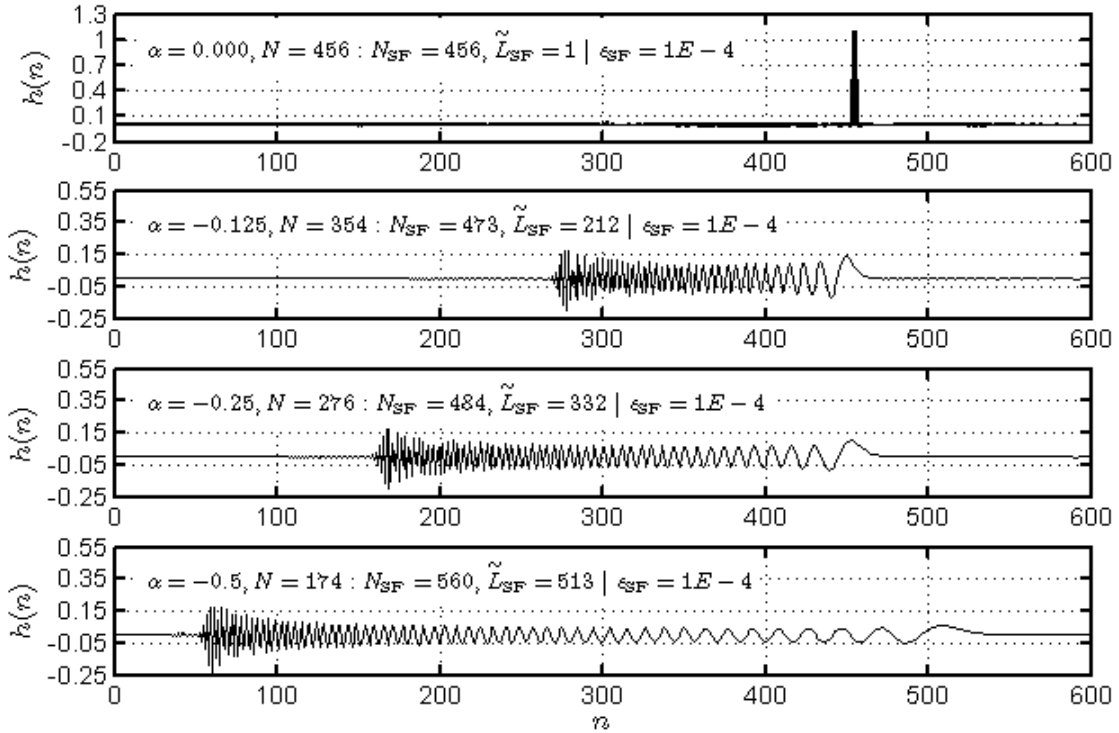


Рис. 15.7 – Импульсные характеристики систем анализ/синтез с параметрами N и α

В работе [6] был предложен метод выравнивания групповой задержки отличающийся от изложенных выше. В [6] показано, что КИХ-фильтр $C(z)$ для выравнивания групповой задержки можно получить путем реверсии импульсной характеристики системы анализ/синтез банка фильтров, для которой требуется корректировка. Даже в случае бесконечного отклика, в нем есть определяющий значимый фрагмент, что проиллюстрировано на рис. 15.7. Следовательно, можно принять, что существуют только отсчеты с индексами между 0 и N_{SF} , которые определяют передаточную функцию полиномиального типа, почти эквивалентную исходной:

$$A^N(z) \approx \sum_{n=0}^{N_{SF}} a(n)z^{-n}. \quad (15.18)$$

Реверсируя порядок следования отсчетов, получаем КИХ-фильтр, который аппроксимирует ряд соединенных обратных передаточных функций и задержек:

$$C(z) = \sum_{n=0}^{N_{SF}} a(N_{SF} - n)z^{-n} \approx A^N(z^{-1})z^{-N_{SF}}. \quad (15.19)$$

Так как $A^N(z^{-1}) = A^{-N}(z)$, то

$$A^N(z)C(z) \approx A^N(z)A^{-N}(z)z^{-N_{SF}} = z^{-N_{SF}}. \quad (15.20)$$

Очевидно, что отклонение произведения стоящего слева от чистой задержки зависит от остатка, который остается после усечений отклика. Нами было экспериментально установлено, что достаточно определить последний значащий отсчет, после которого абсолютные значения становятся меньше $\epsilon_{SF} = 10^{-4}$. Для широких диапазонов α и N , колебания групповой задержки после корректировки с использованием этого порогового значения не превышает нескольких отсчетов для низких частот и быстро уменьшается с увеличением частоты до дробных частей отсчета.

15.2.5 Расчет фильтра-прототипа

Разработано много различных вариантов расчета фильтра-прототипа для косинусно-модулированного банка фильтров (см. [11,12,18] и литературу к этим статьям), большинство из которых включают различные оптимизационные процедуры с нелинейными ограничениями. В данном разделе мы рассмотрим наиболее простой метод расчета фильтра-прототипа описанный в работе [11]. Достоинство данного метода в том, что ограничиваясь поиском фильтра-прототипа среди фильтров, получаемых методом оконного взвешивания, мы сводим задачу проектирования к оптимизации всего одного параметра.

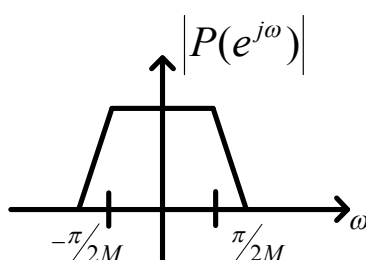


Рис. 15.8 – Типичная АЧХ фильтра-прототипа

На рис. 15.8 схематично изображена АЧХ фильтра-прототипа для косинусно-модулированного банка фильтров.

Допустим, что фильтр-прототип $P(e^{j\omega})$ имеет линейную ФЧХ. Тогда условие почти полного восстановления в банке фильтров может быть выражено в терминах $P(e^{j\omega})$ следующим образом

$$P(e^{j\omega}) \approx 0, \quad |\omega| > \frac{\pi}{2M}, \quad (15.21)$$

$$T(e^{j\omega}) \approx 1, \quad \text{где } T(e^{j\omega}) = \sum_{k=0}^{2M-1} \left| P\left(e^{j\left(\omega - \frac{k\pi}{M}\right)}\right) \right|^2. \quad (15.22)$$

Точность первого приближения (15.21) определяет меру ошибки, возникающую в результате наложения спектров (алайзинга), в то время как точность второго приближения (15.22) задает меру ошибки искажения (от англ. «distortion error»).

Запишем фильтр прототип, получаемый методом оконного взвешивания

$$p(n) = h_i(n)w(n), \quad (15.23)$$

где

$$h_i(n) = \frac{\sin(\omega_c(n - 0.5N))}{\pi(n - 0.5N)}, \quad n = 0, 1 \dots N - 1, \quad (15.24)$$

это импульсная характеристика ФНЧ с частотой среза ω_c , а $w(n)$ — оконная функция (в [11] рассматривался случай, когда $w(n)$ — окно Кайзера). Введем фильтр $G(e^{j\omega}) = |P(e^{j\omega})|^2$, тогда исходя из условия (15.22), можно сделать вывод, что $G(e^{j\omega})$ аппроксимирует фильтр Найквиста ($2M$), это значит, что должно выполняться условие $g(2Mn) \approx (1/2M)\delta(n)$, где $\delta(n)$ — это дельта-функция. Теперь введем простую целевую функцию

$$\phi = \max_{n, n \neq 0} |g(2Mn)|. \quad (15.25)$$

Единственный регулируемый параметр, который влияет на значение функции ϕ это частота среза ω_c . Оптимизация заключается в нахождении значения ω_c , которое бы минимизировало функцию (15.25).

15.3 Объединение субполос

15.3.1 Обоснование

Использование звеньев даже первого порядка, для которых деформация оси частот зависит только от одного коэффициента α , позволяет получить частотное разбиение соответствующее психоакустическим шкалам. Эта проблема была изучена в [18] и выведено выражение в аналитической форме, которое дает коэффициент фазового преобразования для заданной частоты дискретизации.

Рассматриваемый метод с объединением субполос точнее метода, основанного на равнополосном банке фильтров, см. [19], но фактическое его применение к системам с деформированной осью частот не позволяет получить частотное разделение, которое соответствует границам критических полос. Другими словами, точное отображение из линейной оси частот в шкалу барков не обязательно «растягивает» банк фильтров так, что результирующие субполосы соответствуют критическим полосам, особенно, если число последних сравнимо с числом каналов банка фильтров. Хороших результатов можно достичь, если имеется большое число узких субполос, чьи выходы можно интерпретировать как отсчеты в частотной области, и затем просто сгруппировать, чтобы оценить содержимое критической полосы [20].

Критические полосы можно аппроксимировать более точно используя деформацию оси частот на основе фазовых фильтров высших порядков, которые, однако, труднее рассчитывать, применять [21] и реализовывать в аппаратуре.

Другая проблема, связанная с деформацией оси частот, это то, что в предыдущих статьях, коэффициент фазового звена определялся очень точно, в виде десятичной дроби (например, $\alpha = 0.5756$ для частоты дискретизации 16 кГц) при этом предполагалась реализация фазового звена на основе умножителя. По имеющейся у нас информации, никто ещё не рассматривал возможности упростить цепочку фазовых звеньев так, чтобы реализовать её без умножителей.

Мы предлагаем решить проблему аппроксимации критических полос, как по точности, так и по вычислительной сложности путем совмещения деформации оси частот с использованием фазовых фильтров первого порядка с объединением субполос. Идея заключается в проектировании банка фильтров с более либо менее деформированной осью частот, но с числом каналов большим, чем требуется субполос в частотном разбиении, после чего каналы объединяются для аппроксимации критических полос. До сих пор данный подход применялся без особого успеха лишь для равнополосных систем [22–24]. Для систем с деформированной осью частот он представляется более пригодным, как это будет показано в дальнейшем.

15.3.2 Точность частотного разбиения

В нашем подходе при аппроксимации критических полос, главная задача состоит в определении коэффициента фазового преобразования α и необходимого числа каналов так, чтобы число субполос соответствующего банка фильтров после объединения точно соответствовало заданному числу критических полос. Очевиден даже более общий вопрос — как разбить каналы на группы для объединения в субполосы в банке фильтров с деформацией оси частот.

Для решения этой задачи можно объединить субполосы так, чтобы нижняя частота первой из них и высшая частота последней были как можно ближе к краям данной критической полосы. При этом не возникает трудностей, для отдельных субполос, попадающих внутрь аппроксимируемой критической полосы, но возникают затруднения, когда одна субполоса принадлежит двум критическим полосам одновременно. Судя по всему эта задача не имеет удовлетворительного решения, набор расчетных параметров для которых данные ситуации возникают можно просто опускать.

Среди оставшихся пар коэффициентов фазового преобразования и числа каналов можно найти оптимальные комбинации, однако для этого необходимо иметь критерий точности. Мы рассмотрим такие метрики, как максимум и среднеквадратичная ошибка, основываясь на отклонении верхней частоты от заданной. Данное отклонение при этом нормализуют к ширине критической полосы [6]:

$$\epsilon_{\text{MAX}} = \max_{k=0 \dots \tilde{M}-2} \frac{|f_{U_k} - \hat{f}_{U_k}|}{|\hat{f}_{U_k} - \hat{f}_{L_k}|}, \quad (15.26)$$

$$\epsilon_{\text{RMS}} = \sqrt{\frac{1}{\tilde{M}-2} \sum_{k=0}^{\tilde{M}-1} \left(\frac{|f_{U_k} - \hat{f}_{U_k}|}{|\hat{f}_{U_k} - \hat{f}_{L_k}|} \right)^2}. \quad (15.27)$$

В данных выражениях границы k -ой субполосы полученной объединением обозначены как f_{L_k} и f_{U_k} , соответственно, при этом \hat{f}_{L_k} и \hat{f}_{U_k} обозначают соответствующие границы критических полос.

Необходимо подчеркнуть, что во всех выражениях ниже $\tilde{M} \leq M$ обозначает число каналов после объединения.

Имея критерии (15.26) и (15.27), мы можем узнать как они зависят от α и \tilde{M} . Перебор широкого диапазона значений параметров: $-0.9 \leq \alpha \leq 0$ и $20 \leq \tilde{M} \leq 200$, дает поверхность

ошибки. Эксперимент повторялся для 4-х наиболее часто используемых частот дискретизации: 8, 16, 32 и 44.1 кГц, которые соответствуют телефонному диапазону, широкополосной речи, FM и CD качеству соответственно. Полученные поверхности для ошибок максимума показаны на рис. 15.9 с использованием графиков поверхностей.

Чистое объединение (при $\alpha = 0$), позволяет аппроксимировать любое частотное разбиеие с наперед заданной точностью. Однако, при совместном использовании и объединения и деформации оси частот необходимо меньшее число каналов, для получения той же точности.

Для фиксированного числа каналов, с увеличением коэффициента «растяжения» достигается улучшение точности, но лишь до определенной степени: слишком большая деформация оси частот ухудшает аппроксимацию. Если число каналов невелико, то деформация оси частот необходима, чтобы сделать возможным аппроксимацию.

Для фиксированного коэффициента деформации оси частот α , точность всегда может быть улучшена путем добавления каналов. Чем выше частота дискретизации, тем больше каналов необходимо для достижения той же точности, или чтобы сделать возможной аппроксимацию. С другой точки зрения, чем больше разница между самой узкой и самой широкой полосами, которая вызвана расширением частотного диапазона, тем больше деформация оси частот превосходит объединение.

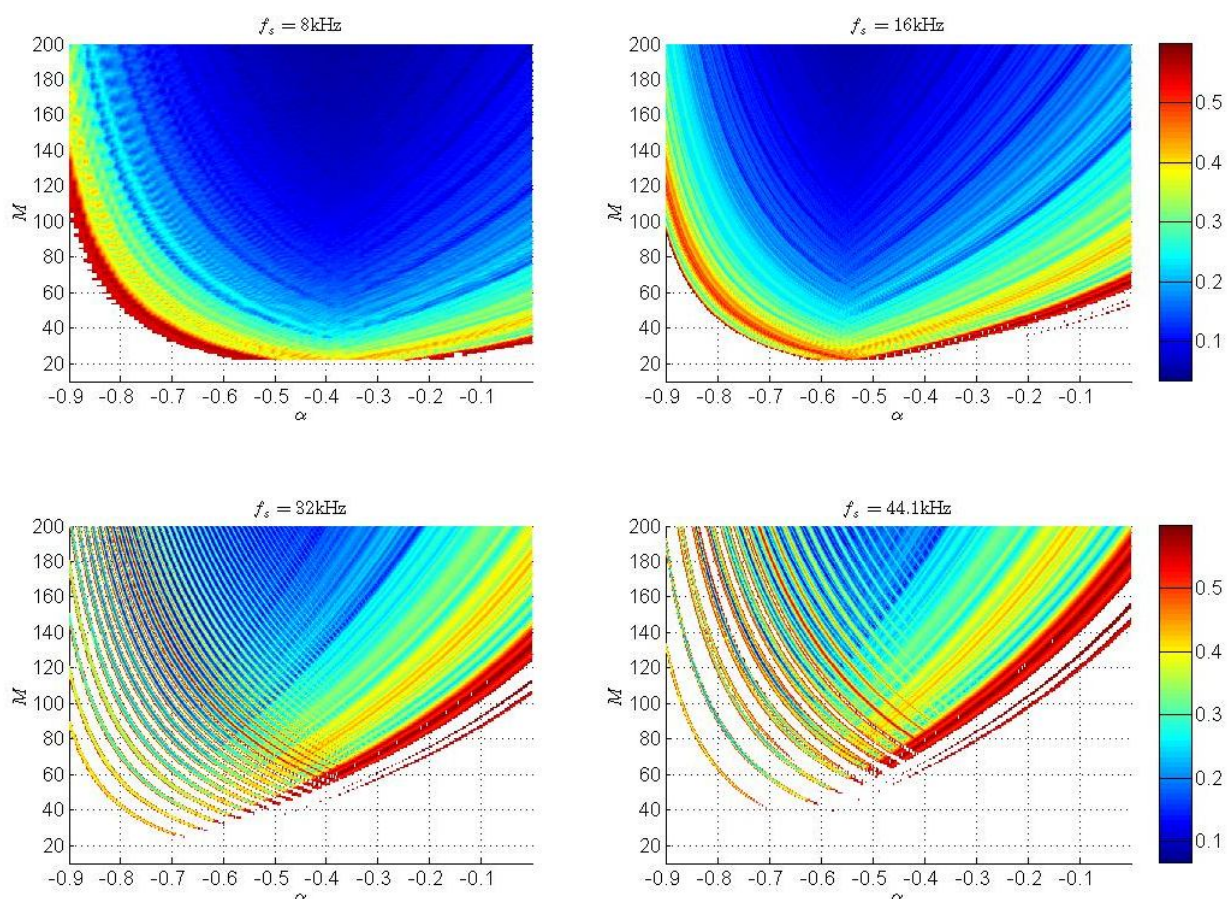


Рис. 15.9 – График профиля максимальной относительной ошибки

Выше были охарактеризованы только общие тенденции. Ошибка аппроксимации колеблется, создавая много изолированных минимумов либо минимумов, формирующих овраги. Из графиков становится понятно, что одна и так же точность может быть получена, используя

различные комбинации коэффициента «растяжения» и числа каналов, которые соответствуют точкам, нерегулярно распределенным по плоскости параметров. Данные сочетания параметров можно использовать для получения оптимальных банков фильтров с точки зрения точности аппроксимации и вычислительной сложности.

15.4 Неравнополосный косинусно-модулированный банк фильтров с частотной деформацией и объединением субполос на основе фазовых звеньев без умножителей

В данном разделе рассматриваются неравнополосные банки фильтров, для которых цепочка фазовых звеньев может быть реализована без использования аппаратного умножителя, т.е. при $\alpha = \{0, -1/2, -1/4, -1/8\}$. Таким образом, применяя теорию из предыдущих разделов, были рассчитаны четыре банка фильтров для аппроксимации критических полос при частоте дискретизации входного сигнала 16 кГц. В табл. 15.2 приведены границы критических полос и информация об объединении каналов.

Табл. 15.2 - Критические полосы и объединения субполос, используемые для их аппроксимации

Критическая полоса		Объединяемые субполосы/Коэффициенты передискретизации S_k для наборов параметров			
№	Частотный диапазон, кГц	$\alpha = -0.5$ $M = 29$	$\alpha = -0.25$ $M = 46$	$\alpha = -0.125$ $M = 59$	$\alpha = 0$ $M = 76$
1	0-100	0 – 0/57	0 – 0/51	0 – 0/50	0 – 0/50
2	100-200	1 – 1/34	1 – 1/31	1 – 1/30	1 – 1/30
3	200-300	2 – 2/24	2 – 2/21	2 – 2/21	2 – 2/21
4	300-400	3 – 3/37	3 – 3/33	3 – 3/33	3 – 3/33
5	400-510	4 – 4/30	4 – 4/27	4 – 4/27	4 – 4/27
6	510-630	5 – 6/22	5 – 5/35	5 – 5/34	5 – 5/35
7	630-770	7 – 7/28	6 – 6/30	6 – 6/30	6 – 6/30
8	770-920	8 – 8/25	7 – 8/15	7 – 8/15	7 – 8/24
9	920-1080	9 – 10/20	9 – 9/28	9 – 9/28	9 – 9/36
10	1080-1270	11 – 11/30	10 – 11/17	10 – 11/17	10 – 11/24
11	1270-1480	12 – 12/27	12 – 13/20	12 – 13/20	12 – 13/20
12	1480-1720	13 – 14/13	14 – 15/17	14 – 15/17	14 – 15/23
13	1720-2000	15 – 15/15	16 – 17/15	16 – 17/15	16 – 18/15
14	2000-2320	16 – 17/9	18 – 19/13	18 – 20/13	19 – 21/13
15	2320-2700	18 – 18/11	20 – 22/11	21 – 23/11	22 – 25/11
16	2700-3150	19 – 20/7	23 – 24/15	24 – 27/7	26 – 29/12
17	3150-3700	21 – 21/8	25 – 28/6	28 – 31/8	30 – 34/8
18	3700-4400	22 – 23/5	29 – 31/7	32 – 36/7	35 – 41/7
19	4400-5300	24 – 24/4	32 – 35/4	37 – 42/4	42 – 49/6
20	5300-6400	25 – 26/2	36 – 39/2	43 – 49/2	50 – 60/2
21	6400-7700	27 – 27/4	40 – 44/4	50 – 56/4	61 – 72/4
22	7700-8000	28 – 28/6	45 – 45/18	57 – 58/18	73 – 75/21

Частотные характеристики банков фильтров до и после объединения каналов показаны на рис. 15.10–рис. 15.13, штриховыми и сплошными линиями соответственно. Дополнительно изображены графики частотной характеристики $|T(f)|$ и групповой задержки $\tau(f)$ функции искажений системы анализ-синтез с выровненной групповой задержкой. Расчетные параметры, порядок фильтров прототипов и значения ошибок аппроксимации показаны на графиках.

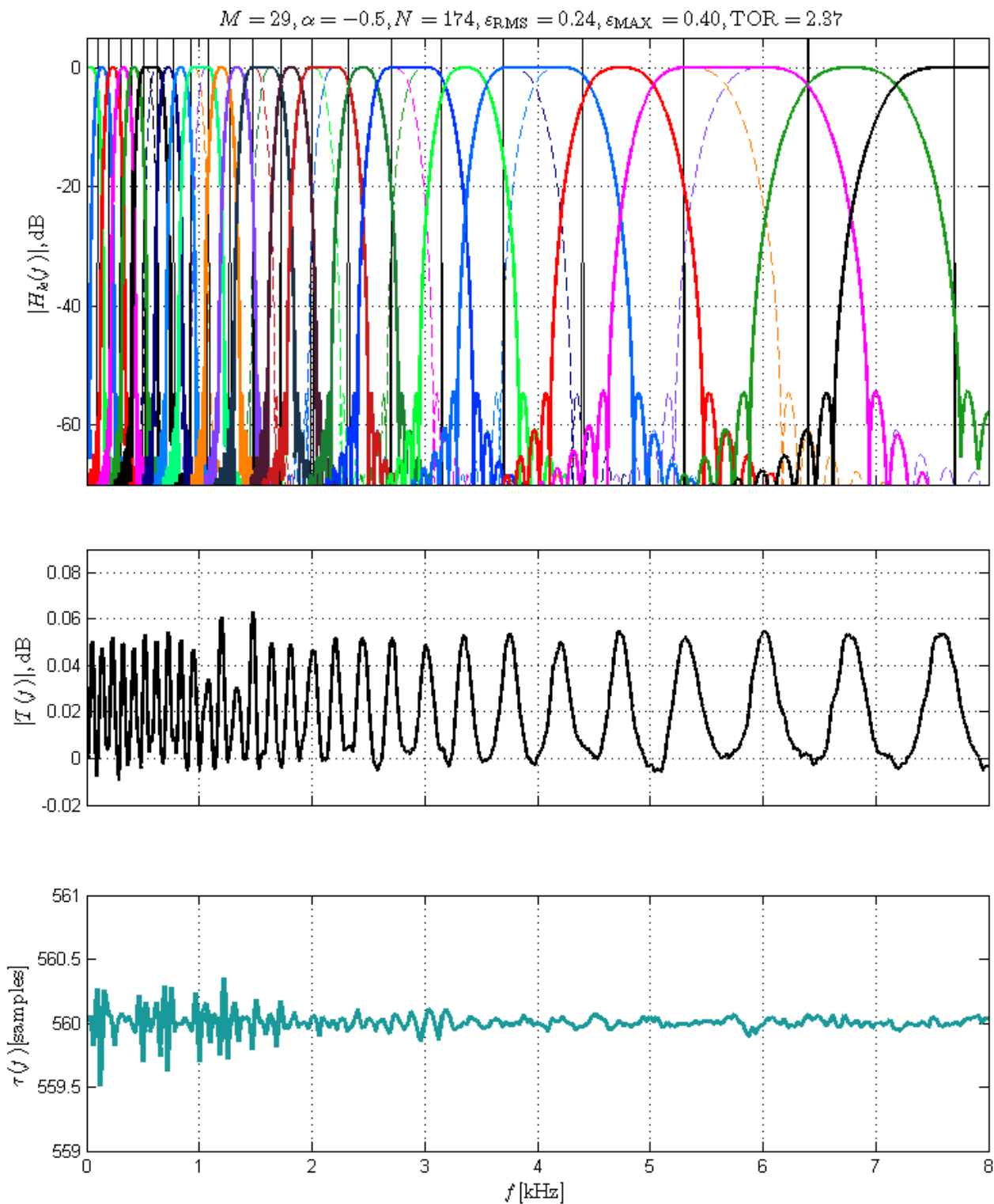


Рис. 15.10 – Частотные характеристики банка фильтров, основанных на деформации оси частот и объединении субполос ($M = 29, N = 174$)

Фильтры-прототипы для банков фильтров были рассчитаны по методу [11], а чтобы ограничить алгоритмическую задержку порядок фильтра N выбирался равным $4M$. Приведенные банки фильтров характеризуются различными компромиссными решениями, между степенью деформации оси частот и числом каналов исходного банка фильтров.

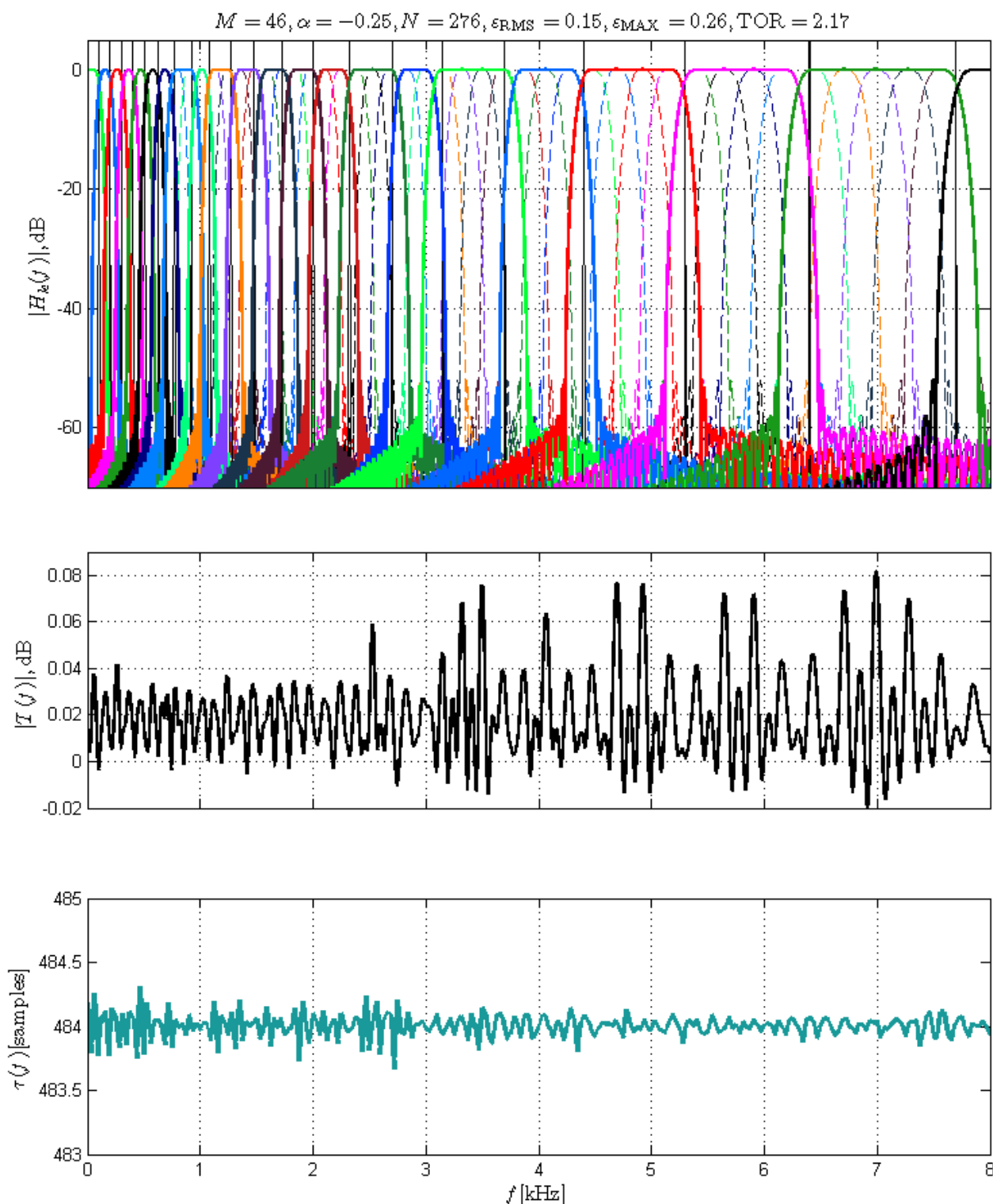


Рис. 15.11 – Частотные характеристики банков фильтров, основанных на деформации оси частот и объединении субполос ($M = 46, N = 276$)

Все банки фильтров, кроме последнего, характеризуются схожей точностью аппроксимации, групповой задержкой и уровнем искажений. Главным отличием между системами является то, как переходная полоса зависит от числа субполос. Чем меньше степень деформации, тем менее она возрастает с увеличением частоты. В двух системах с $\alpha = -1/8$ и $\alpha = 0$ как узкие, так и широкие субполосы характеризуются одинаковой шириной переходной полосы, что является нежелательным во многих приложениях.

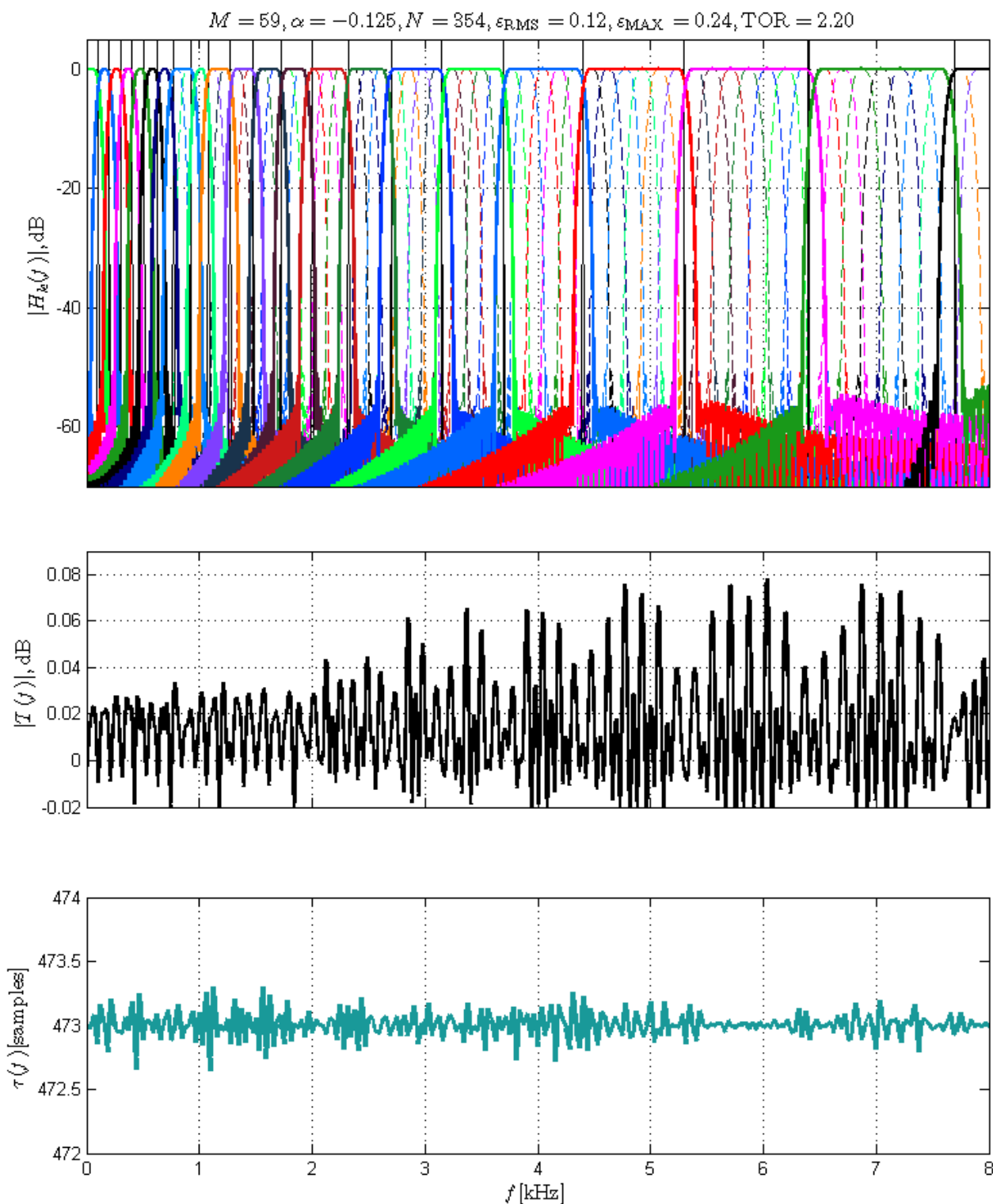


Рис. 15.12 – Частотные характеристики банков фильтров, основанных на деформации оси частот и объединении субполос ($M = 59, N = 354$)

Деформация оси частот позволяет сделать перекрытие между субполосами пропорциональным их ширине полос, так, что длина фильтра-прототипа может быть уменьшена без влияния на уровень ослабления в полосе заграждения. Этот фактор также влияет на общее отношение передискретизации, которое мы обозначим ООП (англ. TOR — total oversampling ratio) и которое вычисляется как сумма величин обратных коэффициентам децимации.

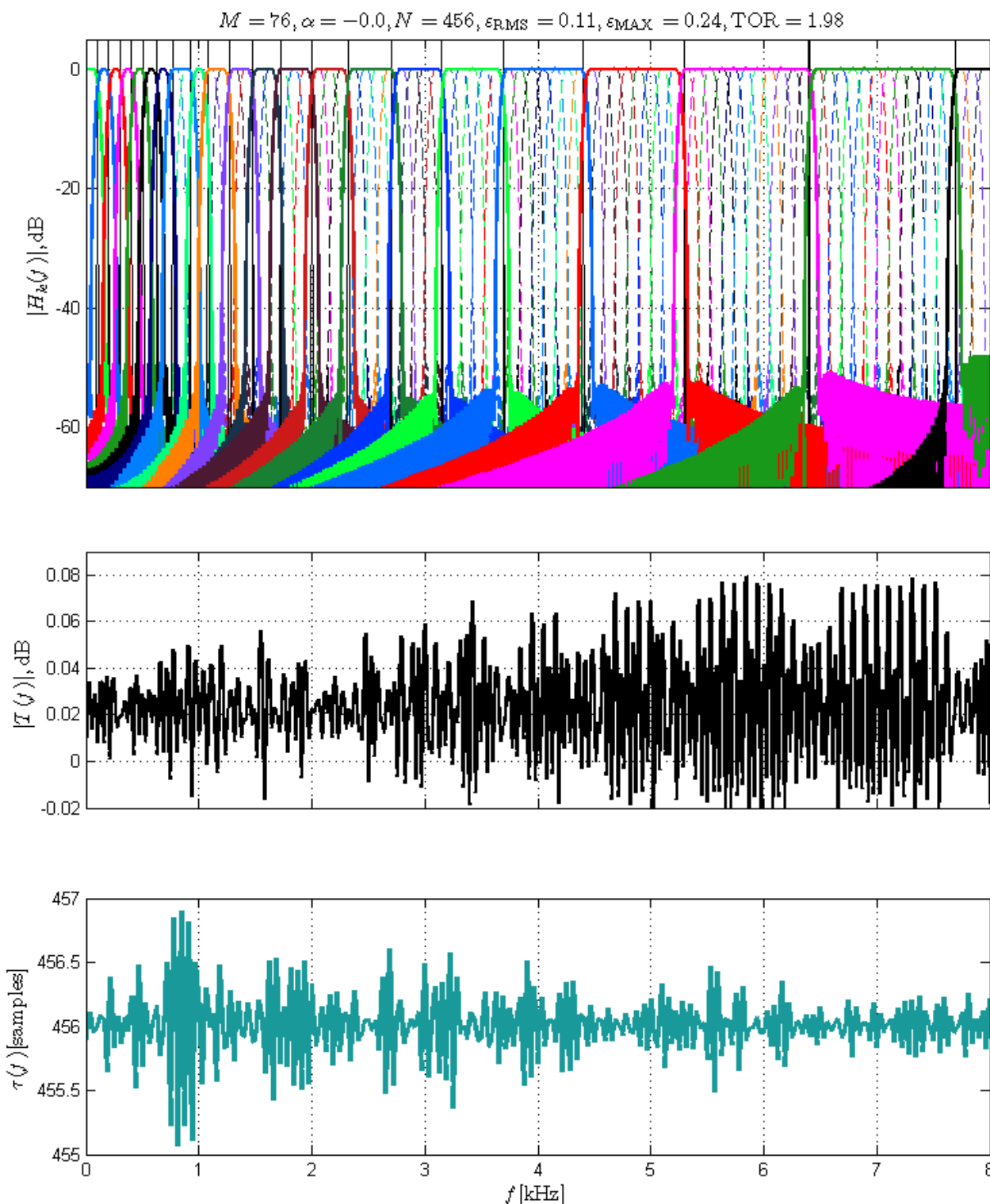


Рис. 15.13 – Частотные характеристики банков фильтров, основанных на деформации оси частот и объединении субполос ($M = 76, N = 456$)

Расширение переходных полос увеличивает риск наложения спектров, так что переходные полосы фильтров должны быть добавлены к краям субполос, перед тем как применять выражение (15.15), что, к сожалению, приводит к более низким коэффициентам децимации.

Рассчитанные банки фильтров характеризуются задержкой в 29-35 мс при ослаблении в полосе задерживания в 50 дБ, что является довольно хорошим результатом и представляется невозможным улучшить его без ущерба и потери качества.

15.5 Список использованных источников

1. Эфрусси М.М., Слуховые аппараты и аудиометры, М.: «Энергия», 1975, 96 с.
2. Marpatin R., Heute U., Antweiler Ch., Advances in digital speech transmission, Chichester.: John Wiley & Sons Ltd, 2008, 573 p.
3. Bäuml R.W., Sörgel W., Uniform polyphase filter banks for use in hearing aids: design and constraints, 16th European signal processing conference (EUSIPCO-2008), Lausanne, Switzerland, 25-29 august 2008.
4. Kurbiel T., Göckler H.G., Alfsmann D., Oversampling complex-modulated digital filter bank pairs suitable for extensive subband-signal amplification, in proc. of 17th European signal processing conference (EUSIPCO-2009), Glasgow, Scotland, 24-28 august 2009, pp. 2658–2662.
5. Alfsmann D., Gockler H.G., Kurbiel T., Filter banks for hearing aids applying subband amplification: a comparison of different specification and design approaches, in proc. of 17th European signal processing conference (EUSIPCO-2009), Scotland, Glasgow, August 24-28, 2009, pp. 2663–2667.
6. Parfieniuk M., Petrovsky A. A., Wan W, Frequency warping and subband merging for approximating the critical bands with cosine modulated filter banks, Internationals conference on audio, language and image processing (ICALIP'08), China, Shanghai, July 7-8, 2008, Vol. 2, pp. 1159–1166.
7. Galijasevic E., Kliewer J., Non-uniform near-perfect-reconstruction oversampled DFT filter banks based on all-pass transform, The Proc. of IEEE Digital Signal Processing Workshop'2000, Hunt, TX, USA, 2000, pp. 1–6.
8. Löllmann H. W., Vary P., Improved design of oversampled allpass transformed DFT filter banks with near-perfect reconstruction, in proc. of 15th European signal processing conference (EUSIPCO-2007), Poland, Pozan, September 3–7,2007, pp. 50–54.
9. Parfieniuk M., Petrovsky A.A., Tunable non-uniform filter bank mixing cosine modulation with perceptual frequency warping by allpass transformation, Automatic Control and Computer Sciences, 2004, Vol. 38, No. 4, pp. 44–52.
10. Piotrowski A., Parfieniuk M., Digital filter banks: analysis, synthesis and implementation for multimedia systems, Wydawnictwo Politechniki Bialostockiej, Bialystok, 2006.

-
11. Lin Y.P., Vaidynathan P.P., A Kaiser window approach for the design of prototype filters of cosine modulated filterbanks, *IEEE Signal Processing Letters*, 1998, Vol. 5, No. 6, pp. 132-134.
 12. Bergen S.W., Antoniou A., An efficient closed-form design method for cosine-modulated filter banks using window function, *Signal Processing*, 2007, Vol. 87, No. 5, pp. 811-823.
 13. Koipillai D., Vaidynathan P.P., Cosine-Modulated FIR Filter Banks Satisfying Perfect Reconstruction" *IEEE Trans. on Signal Processing*, 1992, Vol. 40, No. 4, pp. 770-783.
 14. Parfieniuk M., Petrovsky A.A., Simple rule of selection of subsampling ratios for warped filter banks, *Известия Белорусской Инженерной Академии*, 2003, No.1, pp. 130–134.
 15. Kates M., Arehart K. H., Multichannel dynamic range compression using digital frequency warping, *EUROSIP Journal on applied processing*, 2005, Vol.2005, pp. 3003–3014.
 16. Parfieniuk, M. Petrovsky A.A., Reduced complexity synthesis part of non-uniform near-perfect-reconstruction DFT filter bank based on all-pass transformation, *Proc. of the European conf. on Circuit theory and Design (ECCTD'03)*, Cracow, Poland, 2003, Vol. III, pp. 5–8.
 17. Furtado M.B., Diniz P.S., Netto S.L., Numerically efficient optimal design of cosine-modulated filter banks with peak-constrained least-squares behavior, *IEEE transactions on circuits and systems*, 2006, Vol. 52, No. 3, pp. 597–608.
 18. Smith III, J.O Abel J.S., Bark and ERB bilinear transform, *IEEE Trans. On Speech and Audio Processing*, 1999, Vol. 7, No. 4, pp. 697-708.
 19. Karmakar A., Kumar A., Patney R.K., Design of optimal wavelet packet trees based on auditory perception criterion, *IEEE signal processing letters*, 2007, Vol. 14, No. 4, pp. 240–243.
 20. Parfieniuk M., Petrovsky A.A., Warped DFT as the basis for psychoacoustic model, *Proc. of the International Conference of Acoustic, Speech and Signal Processing (ICASSP'04)*, Moreal, Canada, 2004, Vol. IV, pp.185–188.
 21. Kappellan M., Strauss B., Vary P., Flexible nonuniform filterbanks using allpass transformation of multiple order, in *proc. of 8th European signal processing conference (EUSIPCO-1996)*, Trieste, Italy, 7–13 september 1996, pp. 1745–1748.
 22. Lee J.J., Lee B.G., A design of nonuniform cosine-modulated filter banks, *IEEE Trans. Circuits Syst. II.*, 1995, Vol. 42, No. 8, pp. 2105–2109.
 23. Li J., Nguyen T. Q., Tantaratana S., A simple design method for near-perfect reconstruction nonuniform filter banks, *IEEE Trans. Signal Processing*, 1997, Vol. 45, No.8, pp. 2105–2109.
 24. Niamut O.A., Heusdens R., Subband merging in cosine-modulated filter banks, *IEEE Signal processing letters*, 2003, Vol. 10, No.4, pp. 111–114.

16 МАСШТАБИРУЕМЫЙ КОМПРЕССОР ЗВУКОВЫХ СИГНАЛОВ ДЛЯ IP ТЕХНОЛОГИЙ НА ОСНОВЕ ПЕРЦЕПТУАЛЬНО АДАПТИРУЕМОГО ПАКЕТНОГО ДИСКРЕТНОГО ВЕЙВЛЕТ ПРЕОБРАЗОВАНИЯ (ПДВП)

Петровский Ал.А.

АННОТАЦИЯ

В настоящей главе предлагается перцептуальный кодер звука на основе адаптируемого под сигнал и заданный вычислительный ресурс быстрого вейвлет преобразования с динамической трансформацией частотно-временного плана, базирующегося на подходе динамической трансформации алгоритма (ДТА), вычислении перцептуальной энтропии в области вейвлет коэффициентов, эффективном распределении битов кодирования, учитывающего неидеальность преобразования. Основу системы составляет процессор с динамически реконфигурируемой архитектурой. Вычислительная платформа подсистемы аудиокодера реализована на ЦПОС TMS320C6713 и реконфигурируемом процессоре на основе ПЛИС Xilinx семейства Spartan XC3S2000. В систему масштабируемого компрессора звуковых сигналов включен многоканальный микрофонный вход, что позволяет формировать узконаправленную диаграмму микрофонов.

Ключевые слова – ПДВП, перцептуально адаптируемое ПДВП, IP-технология, звуковой сигнал.

СОДЕРЖАНИЕ

16.1 Введение	353
16.2 Перцептуальный аудиокодер	353

16.3 Динамическая трансформация дерева ПДВП под заданный вычислительный ресурс	355
16.3.1 Временная энтропия вейвлет коэффициентов	355
16.3.2 Динамическая трансформация алгоритма – ДТА-подход.....	356
16.3.3 Оптимальное распределение бит при квантовании вейвлет коэффициентов	360
16.3.4 Структурная схема перцептуального кодера звука.....	363
16.3.5 Организация реконфигурируемой архитектуры процессора на основе ДТА-подхода.....	364
16.4 Масштабируемый компрессор звуковых сигналов для IP технологий	365
16.4.1 Структура системы с мультимикрофонным входом	365
16.4.2 Планарное построение массива микрофонов	366
16.4.3 Структурная схема вычислительной платформы подсистемы аудиокодера.....	370
16.4.4 Технические характеристики масштабируемого компрессора звуковых сигналов.....	371
16.5 Список использованных источников.....	372

16.1 Введение

В последние десятилетия важное место в цифровой обработке аудио сигналов занимают алгоритмы кодирования (компрессии), позволяющие получать компактное представление аудио сигнала в цифровой форме для эффективной передачи по каналам связи или хранения на некотором носителе [1,2]. Именно высококачественное кодирование аудио сигналов становится ключевой технологией в развитии цифровых аудиосистем [1], таких как системы запоминания и редактирования цифровых аудио сигналов на персональных компьютерах (домашние студии), компьютерные мультимедийные системы [3,4], цифровое радиовещание, передача аудио сигналов через узкополосные ISDN каналы и сеть Интернет при проведении телеили видеоконференций [4].

В настоящее время интенсивно развиваются подходы организации архитектур процессоров для решения задач обработки звуковых сигналов в реальном времени, базирующиеся на том, что входной сигнал, как правило, не стационарный и, с точки зрения эффективной его обработки, можно адаптировать алгоритмы и архитектуру процессора под текущий фрейм обработки сигнала. Такие вычислительные системы называются системами с реконфигурируемым процессом обработки сигнала [5], а данные подходы относятся к классу динамических: 1) согласованно со входным сигналом непрерывно (от фрейма к фрейму сигнала) осуществляется адаптация алгоритма обработки; 2) синхронно с трансформацией алгоритма обработки реконфигурируется архитектура процессора.

16.2 Перцептуальный аудиокодер

Структуры большинства кодеров звука на основе психоакустики сходны и могут быть представлены обобщенной схемой (см.рис. 16.1) [1,2]. Входной звуковой сигнал $x(n)$ разбивается на квазистационарные фреймы длительностью от 2 до 50 миллисекунд в зависимости от алгоритмов обработки и методов кодирования. Блок частотно-временного анализа аппроксимирует временные и спектральные параметры звука для каждого анализируемого фрейма с учетом шкалы критических частотных полос. В блоке психоакустического анализа оценивается энергия маскирующего сигнала (уровни маскирующих порогов) на базе психоакустической модели. При этом определяются максимальные искажения, возникающие в каждой точке частотно-временной плоскости в процессе квантования и кодирования частотно-временных оценок без введения искусственного артефакта слышимости при восстановлении сигнала. Следовательно, психоакустический анализатор вычисляет частотно-временной параметр невосприимчивости акустической информации слушателем, который затем передается в блок квантования и кодирования. В процессе психоакустического кодирования необходимо: во-первых, установить вид маскирующего сигнала, во-вторых, вычислить соответствующие пороги, затем полученную информацию использовать для того, чтобы расположить спектр шума кодирования ниже порога маскирования, известный как порог едва различимых искажений JND (just noticeable distortion) [4].

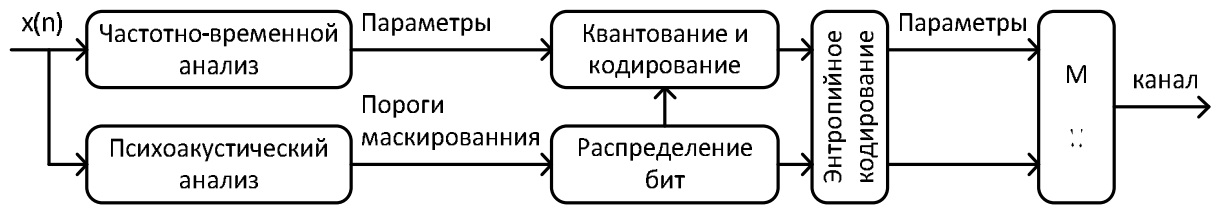


Рис. 16.1 – Общая структура перцептуального кодера звука

Значения психоакустических порогов маскирования неявно показывают результат кодирования звука, а точнее вероятный шаг квантования. Минимальное число бит, необходимое для кодирования воспринимаемого слушателем сигнала, оценивается перцептуальной энтропией PE (perceptual entropy) [6]:

$$PE = \frac{1}{N} \sum_{f=f_l}^{f_h} \max \left(0, \log_2 \frac{|signal(f)|}{threshold(f)} \right), \quad (16.1)$$

где N – число частотных компонент в частотном диапазоне f_l и f_h ; f_l – нижняя частота диапазона; f_h – верхняя частота диапазона, например $f_h = 22050$ Гц; $|signal(f)|$ – амплитуда частотной компоненты f ; $threshold(f)$ – оценка порога маскирования на частоте f . На практике, PE (16.1) часто называют функцией Джонстона и вычисляют на основе субполосного анализа звука [6]:

$$PE = \sum_{i=1}^{25} \sum_{\omega=bl_i}^{bh_i} \log_2 \left(2 \left\lceil \operatorname{nint} \left(\frac{\operatorname{Re}(\omega)}{\sqrt{6 \cdot T_i/k_i}} \right) \right\rceil + 1 \right) + \log_2 \left(2 \left\lceil \operatorname{nint} \left(\frac{\operatorname{Im}(\omega)}{\sqrt{6 \cdot T_i/k_i}} \right) \right\rceil + 1 \right), \quad (16.2)$$

где i – индекс критической полосы; bl_i и bh_i – нижнее и верхнее значение частоты i -ой критической полосы; k_i – количество компонентов преобразования в i -ой критической полосе; T_i – значение порога маскирования в критической полосе i ; $\operatorname{nint}(\)$ – операция округления до ближайшего целого значения. Следовательно, та часть сигнала, которая может быть изменена (в общем случае отброшена) и при этом не вносится дополнительных искажений при его восстановлении, является перцептуально избыточной, а та часть сигнала, отражающая слышимую акустическую информацию человеком, измеряется и кодируется.

Выбор метода частотно-временного анализа неразрывно взаимосвязан с точностью вычисления психоакустической модели и общей концепцией построения системы кодирования сигналов звука. При этом выбор методологии частотно-временного анализа всегда обуславливает решение фундаментального вопроса: обеспечения требуемого разрешения как в частотной так и временной областях. Развитие этих работ в настоящее время идет по пути построения перцептуальных кодеров звука на основе подходящего частотно-временного преобразования, в частности, банков цифровых фильтров и наиболее гибкого, в смысле управления частотно-временным разрешением, быстрого вейвлет преобразования (пакет дискретного вейвлет преобразования (ПДВП)) с динамической трансформацией дерева алгоритма.

ПДВП есть обобщение диадического вейвлет преобразования, которое позволяет получить множество структур путем его соответствующих декомпозиций [7]. Структура ПДВП больше согласуется с сигналом, чем вейвлет преобразование [8], и характеризуется следующи-

ми свойствами: малая вычислительная сложность процедуры декомпозиции звукового сигнала в выбранном базисе (процедура анализа); малая вычислительная сложность процедуры реконструкции сигнала (синтеза) в выбранном базисе; конвейерность вычислительного процесса процедур анализа и синтеза, что способствует организации поточных и параллельно-поточных структур процессоров реального времени; гибкое изменение частотно-временного разрешения, обеспечивающее локализацию нестационарностей в сигнале; единственность преобразования, т.е. в ограниченном числе структур ПДВП имеется одна структура, идентифицирующая соответствующие компоненты сигнала [7].

16.3 Динамическая трансформация дерева ПДВП под заданный вычислительный ресурс

16.3.1 Временная энтропия вейвлет коэффициентов

Меру информативности структуры дерева ПДВП, предлагается конструировать следующим образом [9,10]:

$$H_{E_i} = \sum_{\forall (l,n) \in E_i} \sum_k \frac{|X_{l,n,k}|}{\sum_{\forall (l,n) \in E_i} |X_{l,n,k}|} \ln \left(\frac{|X_{l,n,k}|}{\sum_{\forall (l,n) \in E_i} |X_{l,n,k}|} \right), (l, n) \in E_i, k \in \mathbb{Z}, i = \overline{1,8}, \quad (16.3)$$

где $X_{l,n,k} \in (l, n)$ – коэффициенты узла (l, n) дерева E_i .

Данная стоимостная функция характеризует энтропию вейвлет коэффициентов $X_{l,n,k}$ в узлах (l, n) дерева E_i и отражает изменение во времени информативности ПДВП, отсюда и название временная энтропия вейвлет коэффициентов H_E (wavelet time entropy).

Утверждение. Минимизация стоимостной функции (16.13) ведет к сокращению неопределенности и соответственно к увеличению информативности дерева ПДВП, описывающего входной сигнал.

Доказательство.

Легко видеть, что отношение

$$\frac{|X_{l,n,k}|}{\sum_{\forall (l,n) \in E_i} |X_{l,n,k}|}, \quad (16.4)$$

представляет собой вероятность $p_{l,n}(k)$ k -го вейвлет коэффициента узла (l, n) дерева ПДВП E_i : $p(k)$, $k = 1, 2, 3, \dots$. Пусть частичная сумма определяется следующим образом $S[p](m) \triangleq \sum_{k=1}^m p(k)$, а для упрощения записи полагается, что $S[p](0) = 0$. Если $p(k)$ является функцией вероятности, то $0 \leq S[p](m) \leq 1$ для $\forall m = 1, 2, 3, \dots$. Таким образом, частичные суммы $S[p](m)$ образуют возрастающую последовательность и для $m \rightarrow \infty$ $S[p](m) \rightarrow 1$.

Пусть $p^*(k)$, $k = 1, 2, 3, \dots$ составляют вероятности вейвлет коэффициентов другого дерева ПДВП, у которого, например, $p^*(k)$ представляют собой убывающую последовательность $p^*(k) \geq p^*(k+1)$ для $\forall k$, среднее значение которой также уменьшается, тогда если $q \leq m$, то $qS[p^*](m) \leq mS[p^*](q)$, а если $i \leq q \leq m$, то $(m-i)S[p^*](q) \geq (m-q)S[p^*](i) + (q-i)$

$i)S[p^*](m)$. Таким образом, функция вероятности p более сконцентрированная функция, чем функция вероятности p^* . Следовательно, она имеет меньшую энтропию: $H[p] \leq H[p^*]$.

Так, если $S[p] \geq S[p^*]$, то $1 - S[p] \geq 1 - S[p^*]$ и значение $\log(p^*(q)/p^*(q+1)) \geq 0$ для $\forall q = 1, 2, 3, \dots$. Определяется, что $p^*(0) = 1$, тогда для каждого $q = 0, 1, 2, \dots$, получается, что

$$(1 - S[p](q)) \log \frac{p^*(q)}{p^*(q+1)} \leq (1 - S[p^*](q)) \log \frac{p^*(q)}{p^*(q+1)}. \quad (16.5)$$

Величина $1 - S[p](q) = \sum_{k=q+1}^{\infty} p(k)$, тогда

$$\sum_{q=0}^{\infty} \sum_{k=q+1}^{\infty} p(k) \log \frac{p^*(q)}{p^*(q+1)} \leq \sum_{q=0}^{\infty} \sum_{k=q+1}^{\infty} p^*(k) \log \frac{p^*(j)}{p^*(j+1)}. \quad (16.6)$$

Если изменить порядок суммирования, то получается, что

$$\sum_{k=1}^{\infty} p(k) \sum_{q=0}^{k-1} \log \frac{p^*(q)}{p^*(q+1)} \leq \sum_{k=1}^{\infty} p^*(k) \sum_{q=0}^{k-1} \log \frac{p^*(k)}{p^*(k+1)}. \quad (16.7)$$

Внутренние суммы складываются, тогда

$$\sum_{k=1}^{\infty} p(k) \log \frac{1}{p^*(k)} \leq \sum_{k=1}^{\infty} p^*(k) \log \frac{1}{p^*(k)}. \quad (16.8)$$

Здесь правая часть выражения по определению является энтропией $H[p^*]$ функции вероятности $p^*(k)$. На основании факта, что $\log(t) \leq t - 1$ (равенство соблюдается только при $t = 1$) показывается, что сумма слева больше энтропии $H[p]$ функции вероятности $p(k)$: $\log(p^*(k)/p(k)) \leq (p^*(k)/p(k)) - 1$, а равенство соблюдается только, если $p^*(k) = p(k)$. После умножения на $p(k)$ и суммирования по k получается, что

$$\sum_k p(k) \log \frac{1}{p^*(k)} \leq \sum_k (p^*(k) - p(k)) = 0, \quad (16.9)$$

откуда следует, что

$$\sum_k p(k) \log \frac{1}{p(k)} \leq \sum_k p^*(k) \log \frac{1}{p^*(k)}. \quad (16.10)$$

Таким образом, утверждение доказано.

16.3.2 Динамическая трансформация алгоритма – ДТА-подход

Алгоритм динамической трансформации дерева ПДВП под заданный вычислительный ресурс (параллельно-поточный реконфигурируемый процессор $(m, r_{l,n})$) и мультифреймовую обработку в реальном времени определяется следующим образом [10].

Алгоритм: Динамическая трансформация дерева ПДВП под заданный вычислительный ресурс

Пусть реконфигурируемый вычислительный ресурс $(m, r_{l,n})$ ограничивается величиной C . Предельное дерево ПДВП $CB - WPD: (l, n) \in E_{CB}$. Пусть требуемый вычислительный ресурс i -го дерева ПДВП $(l, n) \in E_i$, определяется величиной c_i , а решение о декомпозиции узла предка (l, n) на узлы потомки $(l + 1, 2n)$ и $(l + 1, 2n + 1)$ будет обозначаться как $split(l, n)$, где l – масштабный уровень преобразования, n – n -й узел на уровне l . Положим, что j является номером фрейма входного аудио сигнала.

Шаг 1. Пусть $j = 1, m = 1, l = 0$ и $split(l, n) = YES, r_{l,n} = YES$, т.е. задан корневой узел $(0,0)$ дерева преобразования – 1-й фрейм входного аудио сигнала, у которого перцептуальная энтропия $PE_{0,0}$ и информативность дерева $H_{E_{0,0}}$, а также реконфигурация процессора разрешена.

Шаг 2. $i = j$, т.е. 1-й фрейм входного сигнала определяет рост дерева ПДВП. Осуществляется декомпозиция сигнала на основе ячейки (1-й ступени процессора) – банка из двух ортонормальных вейвлет фильтров.

Шаг 3. Вычисляются оценки перцептуальной энтропии в узлах декомпозиции и информативности дерева $H_{E_{1,1}}$.

Шаг 4. Проверяется информативность дерева $E_{1,1}$ по сравнению с деревом $E_{0,0}$

ЕСЛИ $H_{E_{0,0}} < H_{E_{1,1}}$,

ТОГДА это не аудио сигнал, параметры не кодировать, переход к шагу 1.

Шаг 5. $l = l + 1$.

ЕСЛИ $l - 1 >$ максимального масштабного уровня предельного дерева $CB - WPD$,

ТОГДА STOP – конец роста дерева $E_{i,m}$ ПДВП.

Шаг 6. Проверить на принадлежность узлов дерева $E_{i,m}$ дереву $CB - WPD: E_{CB}$:

ЕСЛИ $(l, n) \in E_{i,m} = (l, n) \in E_{CB}$,

ТОГДА $r_{l,n} = NO$.

Шаг 7. $m = m + 1$. Рост дерева $E_{i,m}$ осуществляется следующим образом.

Для каждого узла n уровня l :

оценивается и проверяется достаточность вычислительного ресурса дерева $c_{i,m}$:

ЕСЛИ $c_{i,m} > C$,

ТОГДА $r_{l,n} = NO$ и STOP – конец роста дерева $E_{i,m}$ ПДВП.

Выполняется декомпозиция узла предка (l, n) .

Вычисляется перцептуальная энтропия в узлах потомках $PE_{l+1,2n}$ и $PE_{l+1,2n+1}$:

ЕСЛИ $PE_{l,n} \geq PE_{l+1,2n} + PE_{l+1,2n+1}$,

ТОГДА $split(l, n) = YES, r_{l,n} = YES$

ИНАЧЕ $split(l, n) = NO, r_{l,n} = NO$.

Шаг 8. $j = j + 1$. Читать следующий фрейм аудио сигнала, который обрабатывается в соответствии со структурой дерева $E_{i,m}$ ПДВП.

Шаг 9. Оценивается информативность $H_{E_{i,m}}$ дерева $E_{i,m}$:

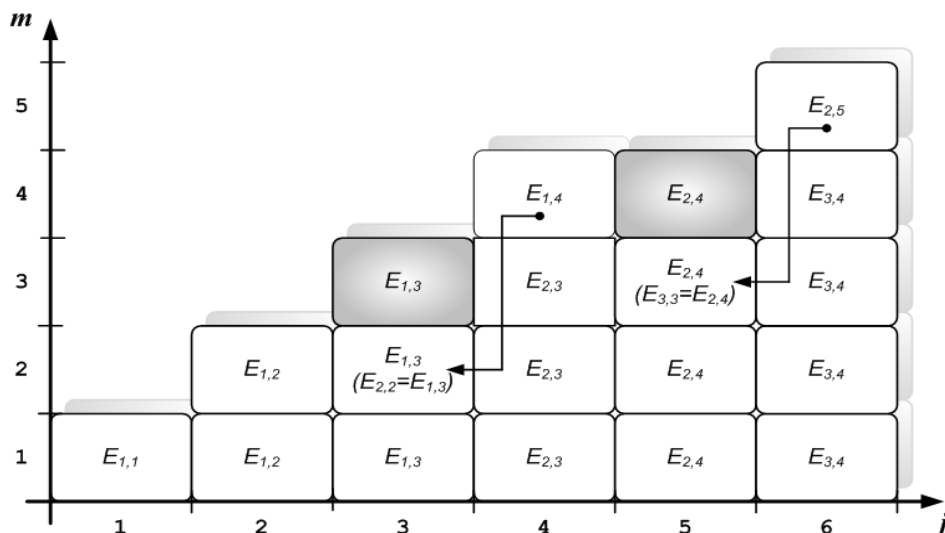
ЕСЛИ $H_{E_{i,m}} > H_{E_{i,m+1}}$,

ТОГДА $r_{l,n} = NO$ и STOP – конец роста дерева $E_{i,m}$ ПДВП.

Шаг 10. Переход к шагу 5.

STOP. Для входного фрейма i оптимальное дерево преобразования $E_{i,m-1}$, $m = m - 2$, $l = l - 2$, $i = i + 1$. Переход к шагу 5.

Пример поточной обработки данных в процессоре, где m – ступень процессора, а j – номер обрабатываемого фрейма, показан на рис. 16.2.



Инициализация: $E_{0,0}$, $l = 0$, $m = 1$, $j = 1$, $i = j = 1$.

1. $l = 1$, $m = 2$, исходное ПДВП дерево $E_{1,1}$, ПДВП дерево потомок $E_{1,2}$, $H_{E_{1,2}} < H_{E_{1,1}}$, $j = 2$
2. $l = 2$, $m = 3$, исходное ПДВП дерево $E_{1,2}$, ПДВП дерево потомок $E_{1,3}$, $H_{E_{1,3}} < H_{E_{1,2}}$, $j = 3$
3. $l = 3$, $m = 4$, исходное ПДВП дерево $E_{1,3}$, ПДВП дерево потомок $E_{1,4}$, $H_{E_{1,4}} > H_{E_{1,3}}$, но

$E_{1,3}$ – оптимальное ПДВП дерево, $m = m - 2 = 2$, $i = i + 1 = 2$, $E_{2,2} = E_{1,3}$, $l = l - 2 = 1$

4. $l = 2$, $m = 3$, исходное ПДВП дерево $E_{2,2}$, ПДВП дерево потомок $E_{2,3}$, $H_{E_{2,3}} < H_{E_{2,2}}$, $j = 4$
5. $l = 3$, $m = 4$, исходное ПДВП дерево $E_{2,3}$, ПДВП дерево потомок $E_{2,4}$, $H_{E_{2,4}} < H_{E_{2,3}}$, $j = 5$
6. $l = 4$, $m = 5$, исходное ПДВП дерево $E_{2,4}$, ПДВП дерево потомок $E_{2,5}$, $H_{E_{2,5}} > H_{E_{2,4}}$, но

$E_{2,4}$ – оптимальное ПДВП дерево, $m = m - 2 = 3$, $i = i + 1 = 3$, $E_{3,3} = E_{2,4}$, $l = l - 2 = 2$

7. $l = 3$, $m = 4$, исходное ПДВП дерево $E_{3,3}$, ПДВП дерево потомок $E_{3,4}$, $H_{E_{3,4}} < H_{E_{3,3}}$, $j = 6$

Рис. 16.2 – Обработка аудиосигнала согласно ДТА-подхода

Достоинством данного алгоритма является то, что рост дерева осуществляется сверху вниз, без возвратов на меньшие масштабные уровни преобразования и необходимости построения полного дерева ПДВП [9]. Данный алгоритм роста дерева ПДВП позволяет определить субоптимальную структуру декомпозиции ПДВП при минимальном числе бит на отсчет звукового сигнала без воспринимаемых на слух искажений, вносимых в процессе кодирования входного сигнала, с другой стороны, меру информативности структуры дерева ПДВП (16.3) в [11,12] рассматривается как оценка скорости передачи, которую можно достичь на соответствующей структуре дерева ПДВП.

На рис. 16.3. На рис. 16.3б для входного сигнала рис. 16.3а показаны этапы работы алгоритма роста дерева ПДВП и формирования частотно-временного плана.

Как видно, по мере увеличения декомпозиции дерева ПДВП порог маскирования – порог едва различимых искажений приближается к такому же порогу (рис. 16.4), который был построен для предельного дерева преобразования *CB – WPD* (рис. 16.5).

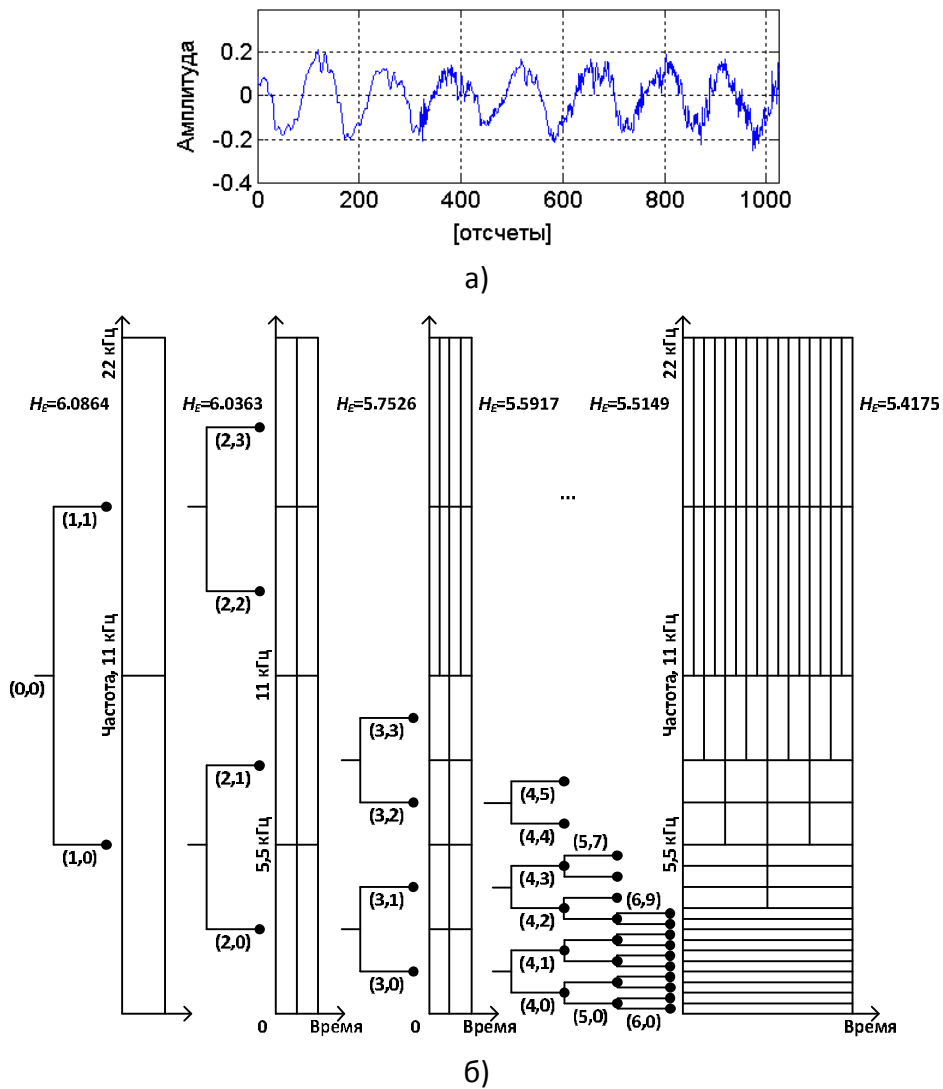


Рис. 16.3 – Формирование дерева ПДВП на основании алгоритма динамической трансформации частотно-временного плана

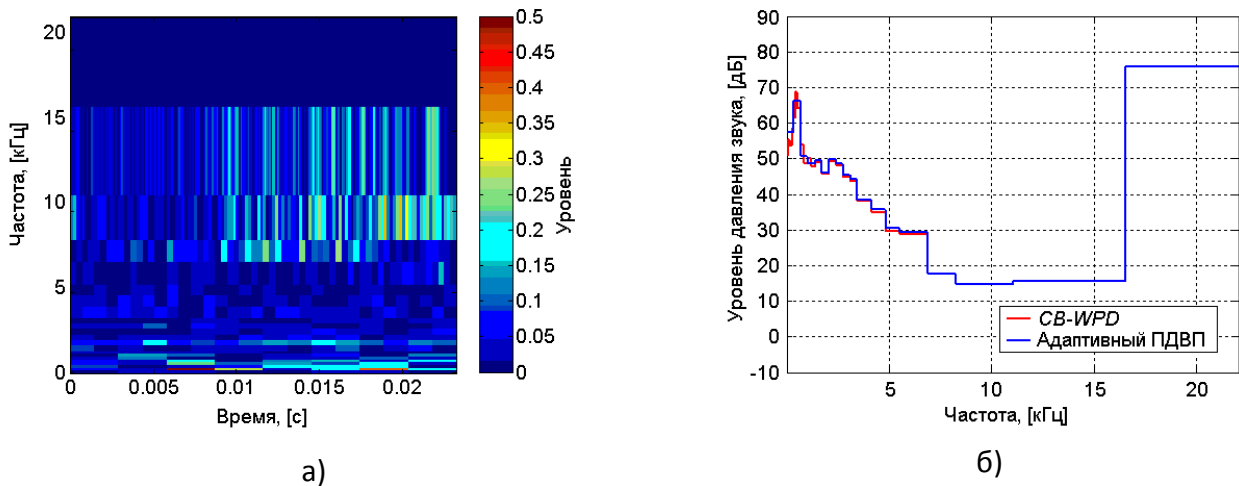


Рис. 16.4 – Спектрограмма и порог маскирования звукового сигнала

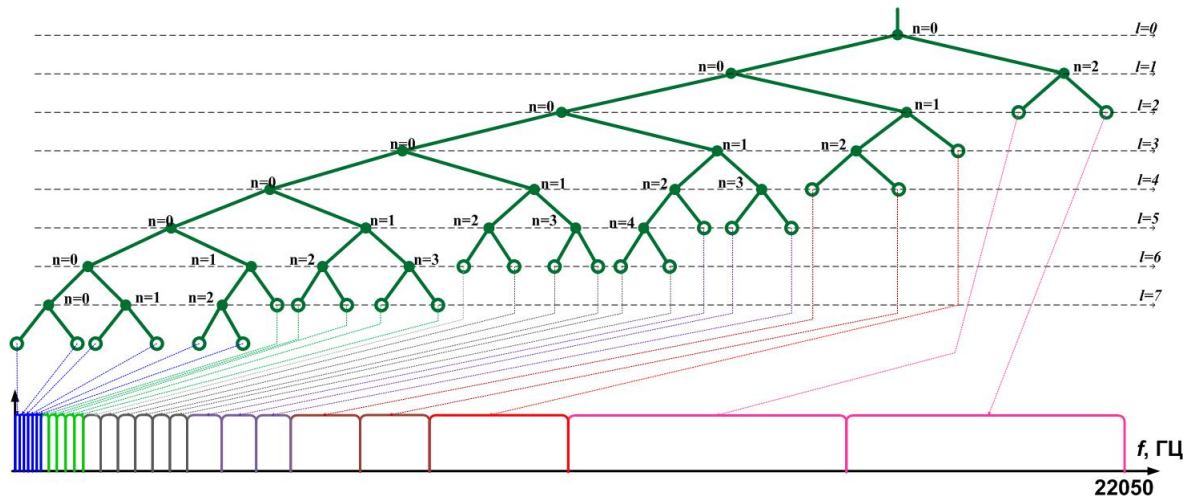


Рис. 16.5 – Дерево ПДВП, аппроксимирующее шкалу барков для частотного диапазона 0-22050 Гц

16.3.3 Оптимальное распределение бит при квантовании вейвлет коэффициентов

Согласно алгоритма роста дерева адаптация шума квантования обеспечивается с помощью перцептуальной модели, в частности, величиной перцептуальной энтропии $PE_{l,n}$ в узле $(l, n) \in E_m$. В [13], предлагается процедура оптимального распределения бит для равнополосного банка анализирующих фильтров. Однако, для декомпозиции частотного диапазона спектра аудио сигнала на неравномерные частотные полосы и при разных коэффициентах децимации в каждой из них, например, как в адаптивном ПДВП, обуславливает разработку соответствующих процедур оптимального распределения бит при квантовании вейвлет коэффициентов $X_{l,n,k}$, $(l, n) \in E_m$, $k \in Z$. При этом, данная процедура должна учитывать не идеальность амплитудно-частотных характеристик (АЧХ) вейвлет фильтров в ПДВП.

Пусть имеется дерево $(l, n) \in E_m$ ПДВП и его вейвлет коэффициенты $X_{l,n,k}$, $1/\mu_{l,n}$ есть коэффициент децимации, $R_{l,n}$ – требуемое число бит в узле (l, n) , т.е. соответствующей частотной полосы, тогда среднее число бит на отсчет входного сигнала следующее

$$R = \sum_{\forall (l,n) \in E_m} \mu_{l,n} \cdot R_{l,n}, \quad (16.11)$$

где для критически дискретизированного банка фильтров [8] $\sum_{\text{для } \forall (l,n) \in E_m} \mu_{l,n} = 1$.

Предполагается, что шум квантования является белым и его спектральная плотность мощности $N(\omega)$ для частотной полосы B входного сигнала, с учётом не идеальности АЧХ фильтров дерева реконструкции сигнала $H_{l,n}$, $(l, n) \in E_m$, равна

$$N(\omega) = \sum_{\forall (l,n) \in E_m} \frac{\sigma_{q_{l,n}}^2}{\mu_{l,n} \cdot B} \cdot |H_{l,n}(\omega)|^2, \quad (16.12)$$

где $\sigma_{q_{l,n}}^2$ - дисперсия внесённых искажений в сигнал для ИКМ квантователя.

Общая мощность шума квантователя равна

$$\sigma_q^2 = \int_B N(\omega) d\omega. \quad (16.13)$$

С другой стороны, когда частотная полоса B входного сигнала с помощью дерева ПДВП E_m разбита на полосы, соответствующие узлам $(l, n) \in E_m$, выражение общей мощности шума квантователя определяется следующим образом:

$$\sigma_q^2 = \sum_i \sum_{\forall (l,n) \in E_m} \int_{B_k} \frac{\sigma_{qi}^2}{\mu_i \cdot B} \cdot |H_i(\omega)|^2 d\omega. \quad (16.14)$$

Соотношение (16.14) может быть представлено в виде:

$$\sigma_q^2 = \sum_i \sigma_{qi}^2 \sum_{\forall (l,n) \in E_m} h_{i,k}, \quad (16.15)$$

где

$$h_{i,k} = \frac{1}{\mu_i \cdot B} \int_{B_k} |H_i(\omega)|^2 d\omega. \quad (16.16)$$

«Коэффициенты просачивания» $h_{i,k}$ энергии из одной полосы в другую, обусловленную не идеальностью АЧХ банка фильтров реконструкции сигнала, задаются квадратной матрицей H , размер которой определяется числом полос разбиения деревом E_m частотной полосы B . Для $i \neq k$ данные коэффициенты представляют собой величину мощности «шума просачивания» от i -го квантователя в k -ю полосу. Если $i = k$, то «коэффициенты просачивания» определяют мощность шума квантователя, обусловленную i -м квантователем в i -ой полосе. Следует заметить, что данная зависимость влияет только на проектирование банка фильтров реконструкции аудио сигнала и нет необходимости корректировки матрицы H в процессе кодирования входного сигнала.

Таким образом, после введения «коэффициентов просачивания» $h_{i,k}$, дисперсия внесенных искажений кодированием, взвешенная маскирующим фактором T_k [14], в реконструированном выходном сигнале перцептуального кодера равна:

$$\sigma_{Tq}^2 = \sum_i \sigma_{qi}^2 \sum_k \frac{h_{i,k}}{T_k} = \sum_i \frac{\sigma_{xi}^2}{3 \cdot 2^{2R_i}} \sum_k \frac{h_{i,k}}{T_k}, \quad (16.17)$$

где σ_{xi}^2 - дисперсия вейвлет коэффициентов i -го узла $(l, n) \in E_m$.

Далее, применяя метод множителей Лагранжа к (16.17), среднее число бит на отсчёт входной последовательности, учитывающее не перфективность частотно-временного преобразования, т.е. ПДВП будет следующее

$$R_i = R + \frac{1}{5} \log_2 \frac{\frac{\sigma_{x_i}^2 \sum_{\forall (l,n) \in E_m} \frac{h_{i,k}}{T_k}}{12 \cdot \mu_i}}{\prod_j \left[\frac{\sigma_{x_i}^2 \sum_{\forall (l,n) \in E_m} \frac{h_{j,k}}{T_k}}{12 \cdot \mu_j} \right]}, \quad (16.18)$$

где $i = 1, 2, \dots$, максимальное значение, которое определяется числом частотных полос соответствующего дерева $(l, n) \in E_m$.

Следует заметить, что выражение (16.18) представляет собой обобщение, полученное для идеальных фильтров. Если $H_i(\omega) = 0 \forall \omega \notin B_i$, то из (16.16) следует, что $h_{i,k} = 0$, если $i \neq k$, и $h_{i,k} = 1$ для $i = k$, а суммы в (16.18) равны $1/T_i$.

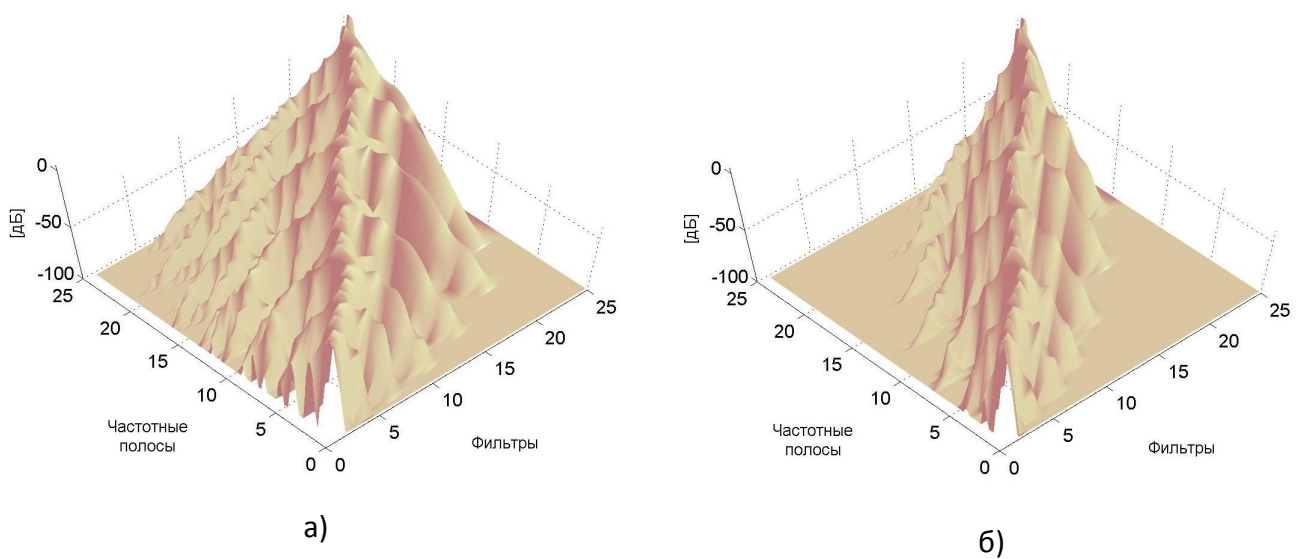


Рис. 16.6 – Графическое представление «коэффициентов просачивания» $h_{i,k}$:

- а) матрица H (25x25), функция Добеши (16 коэффициентов)
- б) матрица H (25x25), функция Добеши (40 коэффициентов)

На рис. 16.6 показано графическое представление «коэффициентов просачивания» для разных порядков вейвлет функций Добеши (8-й рис. 16.6а и 20-й рис. 16.6б) в случае дерева ПДВП $CB - WPD$: 25 полос, частотный диапазон от 0-22050 Гц, матрица H : 25x25. Анализ данных результатов показывает, что эффект просачивания энергии шума квантователя в банке фильтров реконструкции проявляется относительно главной диагонали матрицы H . Более того, наблюдается зависимость «эффекта просачивания» от длины импульсной характеристики фильтров реконструкции сигнала. В частности, для фильтров с короткой импульсной характеристикой данный эффект больше, чем в банке фильтров с длинными импульсными характеристиками.

Таким образом, «эффект просачивания» в смежные полосы энергии шума квантования не может пренебрегаться без внесения ущерба в качество восстановленного аудио сигнала. Выражение (16.18) дает правило оптимального распределения бит при квантовании вейвлет коэффициентов $X_{l,n,k}$ в кодере с учётом эффектов реконструкции сигнала в декодере. Из рис.

16.6б видно, что выбирая достаточно большой порядок вейвлет функций можно ограничиться одинаковыми фильтрами как в анализирующем банке, так и синтезирующем, только если вычислительный ресурс позволяет это сделать.

16.3.4 Структурная схема перцептуального кодера звука

На рис. 16.7 представлена структура перцептуального кодера звукового сигнала, ядром которой является блок «*WPD*», выполняющий реализацию адаптивного ПДВП. В темпе обработки сигналов для каждого текущего дерева E_i вычисляются пороги маскирования $T_{l,n}$, в соответствии с [14], значения перцептуальной энтропии $PE_{l,n}$, а также энтропии H_{E_i} структуры дерева ПДВП E_i . На основании данной информации в блоке «Split» рассчитываются параметры реконфигурации дерева ПДВП $r_{l,n}$ [10]. Данный процесс осуществляется поступательно, без возвратов на меньшие масштабные уровни преобразования. Следовательно, весь вычислительный процесс идеально ложится на архитектуру параллельно-поточных процессоров [15]. Обработка звукового сигнала в ПДВП кодере выполняется кадрами и ввиду того, что преобразование осуществляется над каждым последующим кадром в новой структуре дерева преобразования, то для устранения фазовых разрывов, соседние кадры сигнала перекрываются на 1/8 своей длины. Следующим этапом работы кодера является квантование и кодирование вейвлет коэффициентов $X_{l,n,k}: (l,n) \in E, k \in \mathbb{Z}$, результирующего субоптимального дерева ПДВП E . Управление данным процессом осуществляется в блоке «Алгоритм размещения бит». Кодирование структуры дерева ПДВП E выполняется в блоке «Кодирования структуры дерева ПДВП». Принимая во внимание факт, что рост дерева осуществляется поступательно и изменчивость сигнала во времени более инерционна, чем время обработки, то кодируются только изменения структуры дерева от фрейма к фрейму.

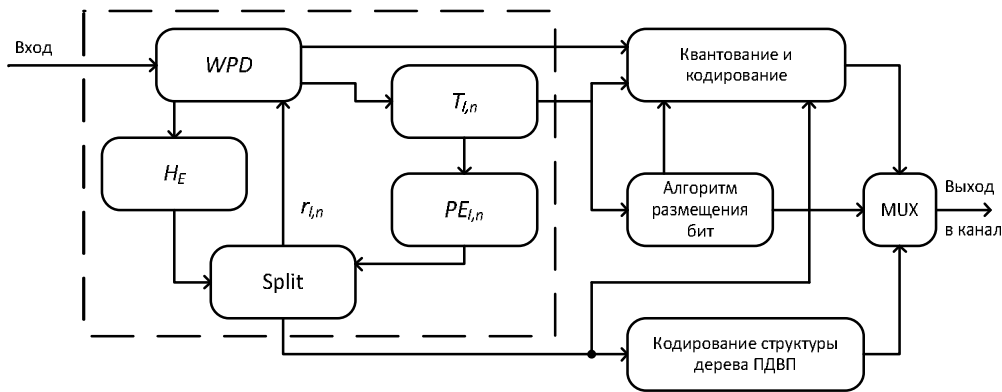


Рис. 16.7 – Структура ПДВП кодера звука

Структура ПДВП декодера звука показана на рис. 16.8. Работа декодера выполняется в следующем порядке: входная информация разделяется на два потока данных в блоке «DEMUX»: один содержит кодированные вейвлет коэффициенты, а второй – код структуры дерева ПДВП; формирование структуры дерева ПДВП происходит в блоке «Формирование структуры дерева ПДВП» и осуществляется путём суммирования кодовой последовательности, описывающей изменение структуры дерева ПДВП, с текущей кодовой последовательностью структуры дерева ПДВП. В случае передачи первого фрейма, структура дерева ПДВП кодируется

полностью в кодере. Реконструкция звукового сигнала выполняется в блоке « WPD^{-1} » на основе обратного адаптивного ПДВП, над декодированными и восстановленными вейвлет коэффициентами (блок «Декодирования и восстановления») в соответствии со структурой дерева ПДВП. Синтезированные кадры реконструированного сигнала умножаются на модифицированное треугольное окно и складываются для формирования выходной последовательности $\hat{x}(n)$.

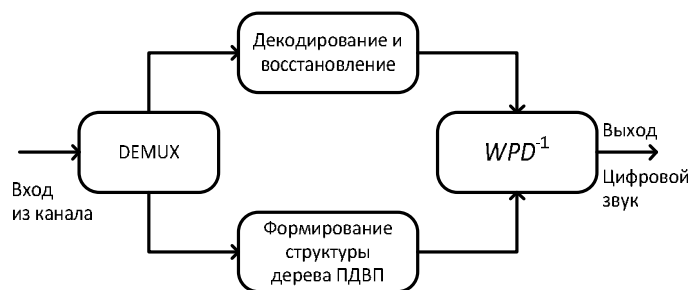


Рис. 16.8 – Структура ПДВП декодера звука

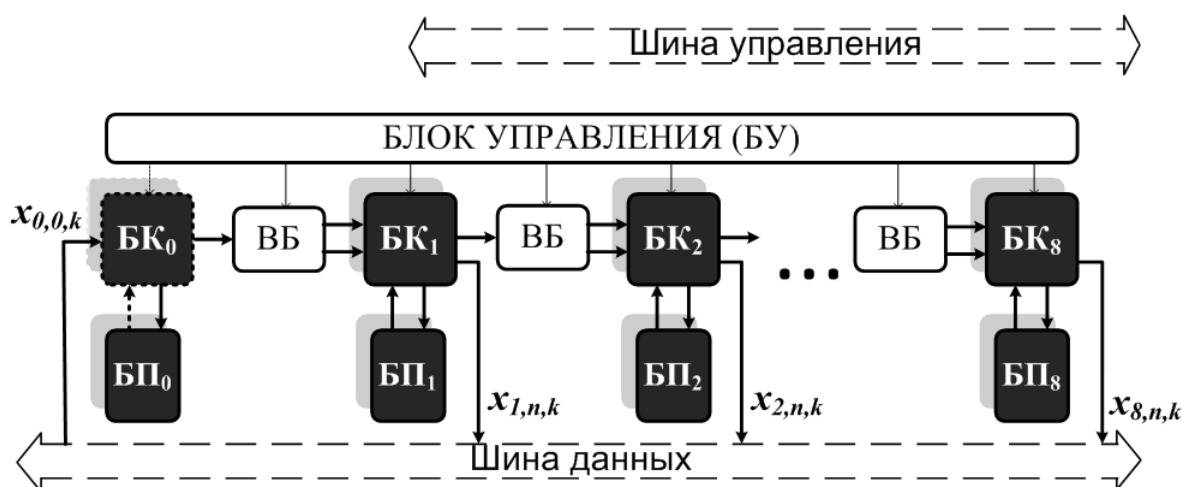


Рис. 16.9 – Структурная схема процессора.

16.3.5 Организация реконфигурируемой архитектуры процессора на основе ДТА-подхода

Структурная организация реконфигурируемой системы ЦОС для аудио кодера на основе подхода динамической трансформации алгоритма (ДТА-подход), предложенного выше, заключается в следующем. Выбирается неоднородная вычислительная система: спецпроцессор, ориентированный на обработку сигнала в частотных полосах, вычисляющий психоакустическую модель и вектор реконфигурации $r_{l,n}$, $(l, n) \in E_i$ процессора ПДВП, и процессор ПДВП с реконфигурируемой архитектурой, который вычисляет ПДВП соответствующего фрейма входного аудио сигнала. Учитывая динамику роста дерева ПДВП, где наблюдается конвейерность и параллелизм в выполнении базовых операций ПДВП – двух вейвлет фильтров низких и высоких частот, а также относительную простоту реконфигурации, выбирается параллельно-поточная архитектура процессора ПДВП, а фильтры реализуются на базе лестничных структур с фиксированной запятой переменного формата [16,17]. Структурная схема процессора показана на рис. 16.9, которая состоит из вычислительных блоков (ВБ), блоков памяти (БП) и блоков коммутации (БК). Блоки памяти и коммутации представлены на рис. 16.10.

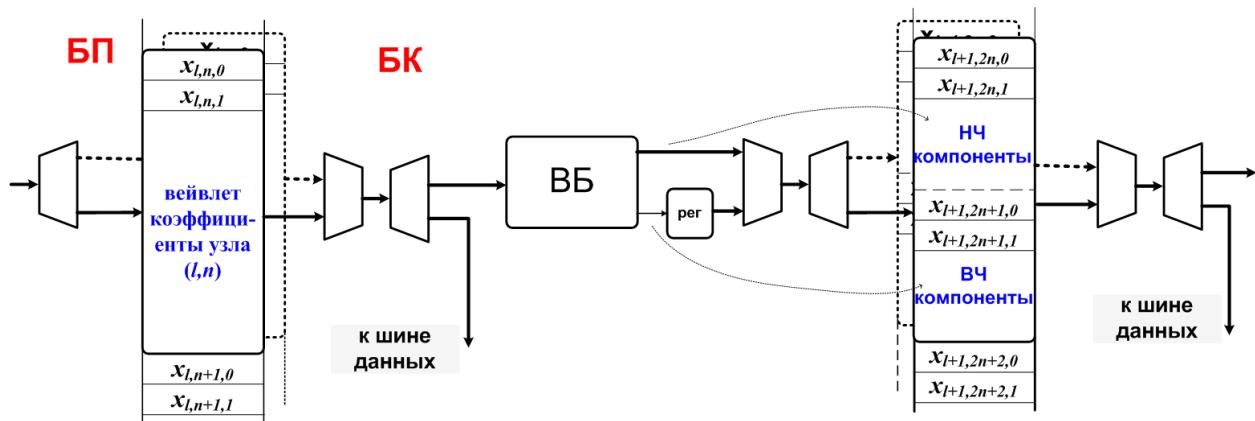


Рис. 16.10 – Блоки памяти и коммутации (БП),(БК).

Вычислительный блок (ВБ) выполняет функцию двухканального банка фильтров, раскладывающего входного сигнала на низкочастотную и высокочастотную составляющие с последующей децимацией по основанию 2 в каждом выходном канале.

16.4 Масштабируемый компрессор звуковых сигналов для IP технологий

16.4.1 Структура системы с мультимикрофонным входом

Структура аудио кодера реального времени мобильной мультимедийной системы состоит из двух подсистем (рис. 16.11): подсистемы аудио кодера, организованной на ДТА-подходе; подсистемы ввода сигнала на основе массива микрофонов, которые организуют супернаправленный микрофон. Синхронизацию данных подсистем и взаимодействие с каналом связи, осуществляет host-процессор – микроконтроллер.

Подсистема аудио кодера состоит из двух ПДВП-процессоров с реконфигурируемой архитектурой: первый процессор – кодировщик, архитектура которого меняется в соответствии с вектором трансформации $r_{l,n}$, формируемым ЦОС-процессором системы мониторинга ДТА-подхода; второй процессор – декодировщик, конфигурация архитектуры которого задается в кодовом слове $code(X_{l,n,k}, (l,n) \in E_i, k \in \mathbb{Z})$, принятым из канала связи. Структура вычислительной платформы кодера звука на основе ДТА-подхода включает управляющий процессор цифровой обработки сигналов (ЦПОС) и ПДВП процессор с реконфигурируемой архитектурой. На ЦПОС возложены следующие задачи: обработка вейвлет коэффициентов $X_{l,n,k}$ в частотных полосах (l,n) , соответствующих текущей структуре дерева E_i ; вычисление психоакустической модели дерева $(l,n) \in E_i: T_{l,n}$ и $PE_{l,n}$; получение вектора реконфигурации ПДВП процессора $r_{l,n}, (l,n) \in E_i$; формирование и разборка данных передачи $[code(X_{l,n,k}, (l,n) \in E_i, k \in \mathbb{Z})]$ и приема $\hat{X}_{l,n,k}(code), (l,n) \in E_i, k \in \mathbb{Z}$, если используется дуплексный режим работы кодера звука. ПДВП процессор реализуется на параллельно-поточной архитектуре с динамической реконфигурацией. Выбор данной архитектуры обусловлен алгоритмом роста дерева ПДВП. Процессор работает в дуплексном режиме: конвейер кодера $X_{l,n,k} = WPD_{l,n}(x_j(t)), (l,n) \in E_i$, $x_j(t)$ – j -й фрейм звукового сигнала и конвейер декодера $\hat{x}_j(t) = WPD_{l,n}^{-1}(\hat{X}_{l,n,k}), (l,n) \in E_i$,

$\hat{x}_j(t)$ – j -й фрейм реконструированного звукового сигнала. Длина конвейера параллельно-поточного процессора определяется предельной структурой дерева ПДВП ($CB - WPD$), см. [10,15]. Анализ ёмкостной сложности реализации базовой операции ПДВП (пара КИХ фильтров нижних и верхних частот с децимацией), модулей коммутации и буферизации, требуемый аппаратный ресурс ПДВП процессора следующий: объём КЛБ для построения 8-ми модулей базовой операции ПДВП на MAC архитектуре – 2048, и для схем буферизации и переключения – 504. Расчетное число КЛБ с учётом 20% надбавки, обусловленной накладными расходами при реализации на конкретной ПЛИС, равно 3062 КЛБ.

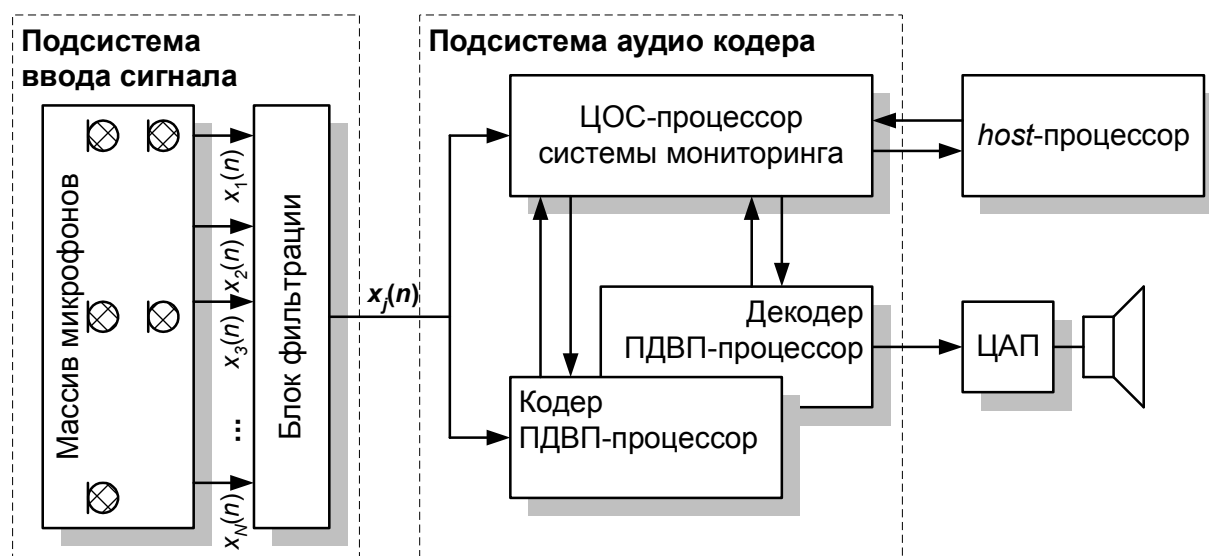


Рис. 16.11 – Структура аудио кодера реального времени мобильной мультимедийной системы

Подсистема ввода сигналов с мультимикрофонным входом обеспечивает приём и оцифровку аудио сигнала. Для получения малого значения SNR необходимо очень близкое расположение микрофона к источнику сигнала, или он должен быть сверхнаправленным. Однако, применение массива недорогих микрофонов со сфокусированной диаграммой направленности позволяет адаптироваться к направлению источника сигнала по максимальной чувствительности, например, Generalized Side-lobe Canceller [18], и может автоматически отслеживать его позицию.

16.4.2 Планарное построение массива микрофонов

Анализ решений на основе подхода массива микрофонов [18,19] показывает, что планарная позиция микрофонов в плоскости более предпочтительнее, чем линейная позиция микрофонов в плоскости [18]: возможно управление чувствительностью в двух направлениях, а так же, как будет показано ниже, усиление сигнала значительно больше.

На рис. 16.12 показана 4-х микрофонная подсистема ввода: микрофоны расположены в плоскости (X, Y) с соответствующими цифровыми фильтрами с импульсными $a_i(n)$, и частотными $A_i(\Omega)$ характеристиками: $i = \overline{1,4}$; $\Omega = 2\pi f/f_s$ – нормированная частота, f_s – частота дискретизации, (X, Y, Z) – декартова система координат. Банк фильтров необходим для формирования результирующего сигнала $x(n)$. Векторы \vec{r}_i определяют позицию i -го микрофона относительно начала системы координат, $x_i(n)$ его выходной сигнал.

Направление на источник сигнала задаётся вектором \vec{u}_0 , который определяется углами наклона θ_0 и поворота ϕ_0 (азимут) плоскости микрофонов. Основываясь на допущении поля в дальней зоне задержка сигнала источника к расположению микрофонов по отношению к опорному сигналу $x_0(n)$ в первоначальной системе координат может быть вычислена на основе планарной модели волны, а на допущении поля в ближней зоне можно применить сферическую модель волны и рассчитать ослабление волны от источника сигнала к каждому микрофону.

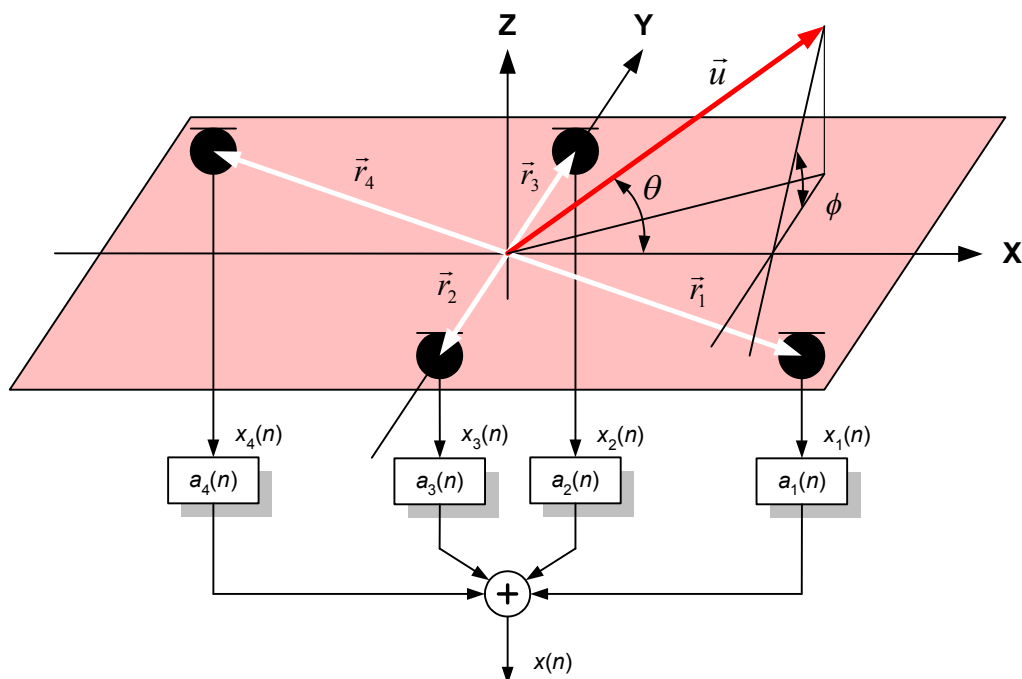


Рис. 16.12 – Планарное построение массива на базе четырех микрофонов

В общем случае, характеристика направленности массива микрофонов может быть описана при помощи диаграммы направленности $\Psi(\Omega, \theta, \phi)$, которая определяется через нормализованную спектральную плотность входного сигнала $x(n)$ как функция от частоты Ω , вектора направленности, т.е. углов наклона θ и азимута ϕ . Используя допущение поля в дальней зоне диаграмма направленности может быть записана как

$$\Psi(\Omega, \theta, \phi) = \left| \sum_{n=1}^N A_n(\Omega) \exp(j\beta \cdot \vec{r}_i^T \vec{u}) \right|^2, \quad (16.19)$$

где $\beta = 2\pi f/c = \Omega f_s/c$ – номер волны, c – скорость звука.

Опираясь на графические обозначения, показанные на рис. 16.12, скалярное произведение $\vec{r}_i^T \vec{u}$ в (16.19) вычисляется следующим образом

$$\vec{r}_i^T \vec{u} = r_{xi} \cos \theta + r_{yi} \cos \phi \sin \theta, \quad (16.20)$$

где r_{xi} и r_{yi} – координаты X и Y вектора \vec{r}_i , соответственно.

Для $r_{yi} = 1, i = \overline{1, N}$, массив микрофонов является линейным и, следовательно, диаграмма направленности не зависит от угла поворота ϕ .

Направленность измеряется через вычисление значения коэффициента усиления $G(\Omega)$, который может быть определен как разность между диаграммами направленности для основного направления падения вектора \vec{u}_0 , $\Psi(\Omega, \theta_0, \phi_0)$ и средним значением направленности относительно всех направлений вектора \vec{u} ,

$$G(\Omega) = \frac{\Psi(\Omega, \theta_0, \phi_0)}{\sum_{i=1}^N \sum_{m=1}^N A_i(\Omega) A^*(\Omega) h_{mi}(\Omega)}, \quad (16.21)$$

где $h_{mi}(\Omega)$ – нормализованная взаимная спектральная плотность мощности сигналов от микрофонов m и i .

Для сферически однородного шумового поля $h_{mi}(f)$ определяется как

$$h_{mi}(\Omega) = \begin{cases} \frac{\sin(\Omega f_s d_{im}/c)}{\Omega f_s d_{im}/c} & \text{при } i \neq m, \\ 1 & \text{при } i = m \end{cases}, \quad (16.22)$$

где $d_{im} = \|\vec{r}_m - \vec{r}_i\|$ - определяет расстояние между i -ым и m -ым микрофоном.

Когда шум в сигнале микрофона является пространственно не коррелированным, то коэффициент усиления массива микрофонов может быть рассчитан на основе коэффициента усиления белого шума:

$$G_W(\Omega) = \frac{\Psi(\Omega, \theta_0, \phi_0)}{\sum_{i=1}^N \sum_{m=1}^N |A_i(\Omega)|^2 h_{mi}(\Omega)}, \quad (16.23)$$

Инверсное значение $G_W(\Omega)$ называется восприимчивостью массива микрофонов $K(\Omega) = 1/G_W(\Omega)$. Восприимчивость массива микрофонов является основным параметром, по которому рассчитывается его чувствительность. Возможные ошибки – это собственный шум, ошибка фазы и амплитуды каждого микрофона массива.

Таким образом, цель проектирования супернаправленного массива микрофонов заключается в выборе коэффициента усиления как можно большим, но так, чтобы значение параметра восприимчивости не превосходило заранее установленного верхнего предела. Передаточная функция $A_i(\Omega)$ i -го фильтра определяется как,

$$A_i(\Omega) = \frac{\tilde{A}_i(\Omega)}{\sum_{m=1}^N \tilde{A}_m(\Omega) \exp(j\beta \cdot \vec{r}_i^T \vec{u}_0)}, \quad (16.24)$$

где $\tilde{A}_i(\Omega)$ - является результатом решения системы линейных уравнений

$$\sum_{m=1}^N h_{im}(\Omega) \tilde{A}_m(\Omega) + \mu \tilde{A}_m(\Omega) = \exp(j\beta \cdot \vec{r}_i^T \vec{u}_0), \quad i = \overline{1, N}, \quad (16.25)$$

В (16.25) μ - множитель Лагранжа, с помощью которого контролируется сверхнаправленность, а также и восприимчивость массива микрофонов. Для сверхнаправленной диаграммы направленности множитель Лагранжа выбирается из интервала $0 \leq \mu \leq 1$. Направленность и восприимчивость массива микрофонов возрастает, когда значение μ стремится к нулю. Значение множителя Лагранжа μ не может быть получено непосредственно из заранее заданной величины восприимчивости $K(\Omega)$. Для этого используется интерактивная процедура поиска оп-

тимального значения μ для каждой частоты f [19]. Импульсные характеристики $a_i(k)$ получаются после решения системы уравнений (16.24) и (16.25) для эквидистантных частот $0 \leq \Omega_i \leq \pi$ путём вычисления обратного дискретного преобразования Фурье длиной M над частотной характеристикой $A_i(\Omega)$ и умножения на окно Хемминга.

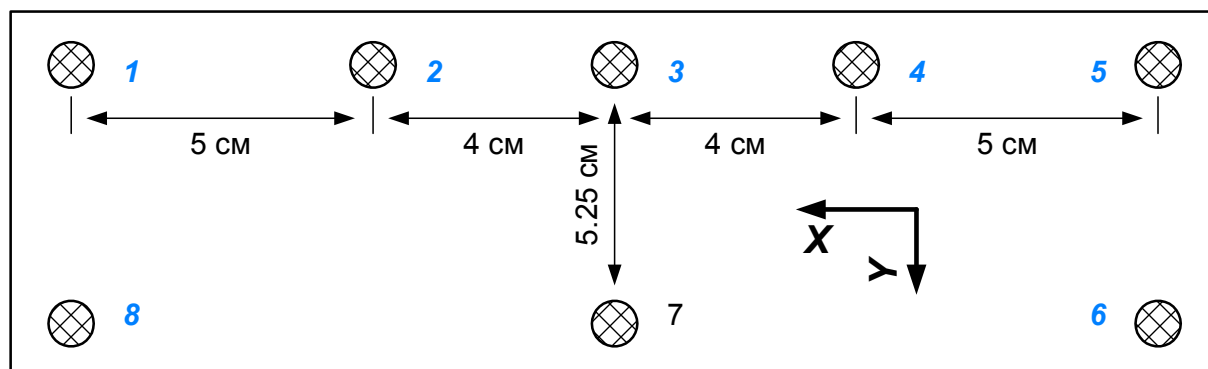


Рис. 16.13 – Экспериментальное расположение микрофонов

На рис. 16.13 показано экспериментальное расположение массива микрофонов. Было найдено, что использование более пяти сенсоров из имеющихся восьми не дало значительного преимущества. Ориентация плоскости микрофонов (рис. 16.12) следующая: ось X направлена к источнику сигнала, углы наклона $\theta_0 = 70^\circ$ и азимута $\phi_0 = 102^\circ$, расстояние от массива микрофонов до источника сигнала приблизительно 40 см, тип микрофона Primo EMU 4747.

Табл. 16.1 – Лучшая линейная и планарная конфигурации массива из пяти микрофонов

Конфигурация	Тип формируемого луча	Номера микрофонов
G5L-DS	Delay-and-sum	1,2,3,4,5
G5P-DS	Dealy-and-sum	1,2,4,5,7
G5L	Superdiective	1,2,3,4,5
G5P	Superdiective	1,2,4,5,7

Лучшая линейная и планарная конфигурации массива из пяти микрофонов в соответствии с методом delay-and-sum формирования диаграмм направленности (см. рис. 16.14), приведены в табл. 16.1.

Оценка различных конфигураций выполняется по значениям коэффициентов усиления и диаграмм направленности. Восприимчивость $K(\Omega)$ проектируемого массива была равна 2. Сравнение коэффициента усиления массива из пяти микрофонов для двух различных типов формирования диаграммы направленности: супернаправленный (G5L и G5P) и delay-and-sum (G5L-DS и G5P-DS) показано на рис. 16.15. Вектор направления на источник сигнала задаётся следующими углами азимута $\phi_0 = 102^\circ$ и наклона $\theta_0 = 70^\circ$. Сравнение линейной конфигурации микрофонов с вариантом delay-and-sum формирования диаграммы направленности с супернаправленной показывает, что супернаправленный подход улучшает результат максимум на 2 дБ и вариант конфигурации G5P (см. табл. 16.1) имеет коэффициент усиления больше, чем вариант конфигурации G5L во всей области частот.

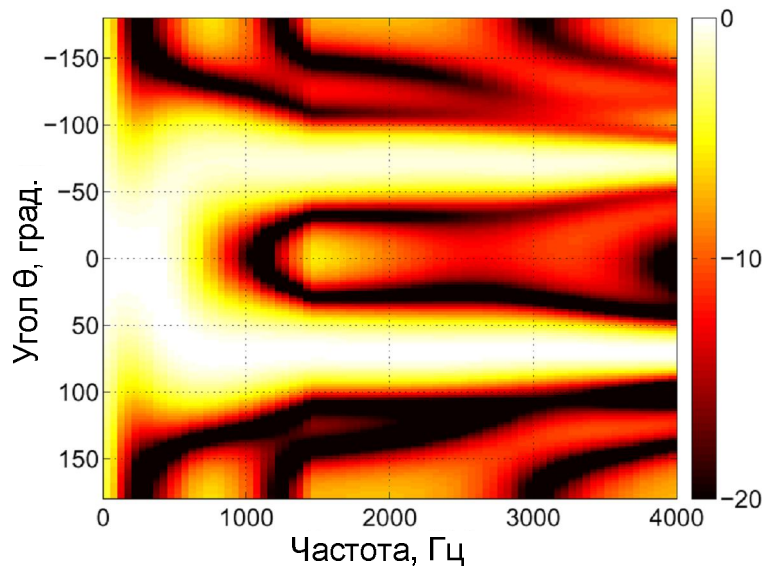


Рис. 16.14 – Диаграмма направленности G5P для массива из пяти микрофонов

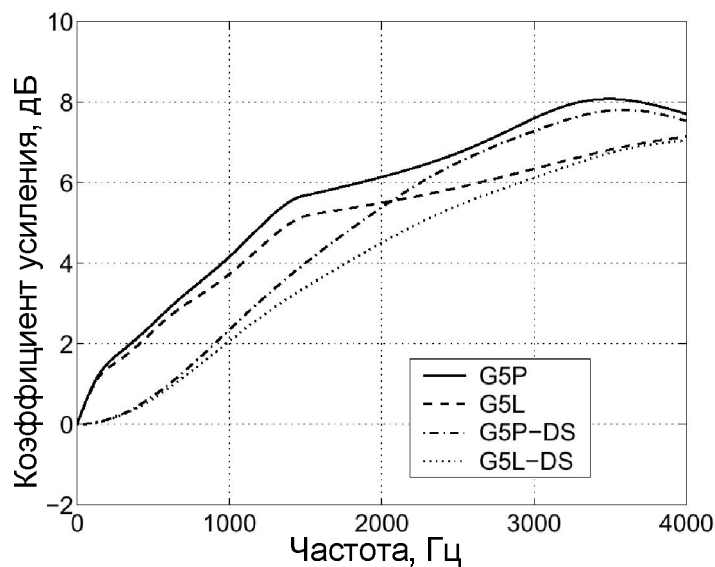


Рис. 16.15 – Коэффициент усиления 1) delay-and-sum массивов G5L-DS (пунктирная) и G5P-DS (штрих пунктирная); 2) супернаправленных массивов G5L (штриховая) и G5P (сплошная).

16.4.3 Структурная схема вычислительной платформы подсистемы аудиокодера

Структурная схема вычислительной платформы представлена на рис. 16.16 [20]. Управляющий процессор реализован на ЦПОС TMS320C6713, а ПДВП процессор синтезирован на ПЛИС Xilinx семейства Spartan-3 XC3S2000 со структурой FPGA. На ПЛИС реализован также внешний цифровой интерфейс аппаратной платформы. Внешняя память SDRAM общим объемом 8 Мбайт и ПЛИС подключаются к TMS320C6713 через 32-разрядный интерфейс внешней памяти EMIF. Интерфейс внешней памяти EMIF имеет четыре отдельные адресные области CE0-CE3. Flash память общим объемом 2 Мбайта предназначена для хранения программ. Процессор использует Flash память в загрузочном режиме. После начальной инициализации модуля программа из Flash памяти автоматически с помощью начального загрузчика загружается во

внутреннюю память управляющего процессора. Для работы со стерео звуком используется кодек TLC320AIC23. Кодек TLC320AIC23 подключается к ЦПОС с помощью двух последовательных каналов: McBSP0 – для доступа к конфигурационным регистрам, McBSP1 – для приема и передачи данных. Период частоты синхронизации может настраиваться за счет использования блоков DMC (Digital Clock Manager) ПЛИС. Для реализации внешнего интерфейса системы используются порты ввода/вывода ПЛИС.

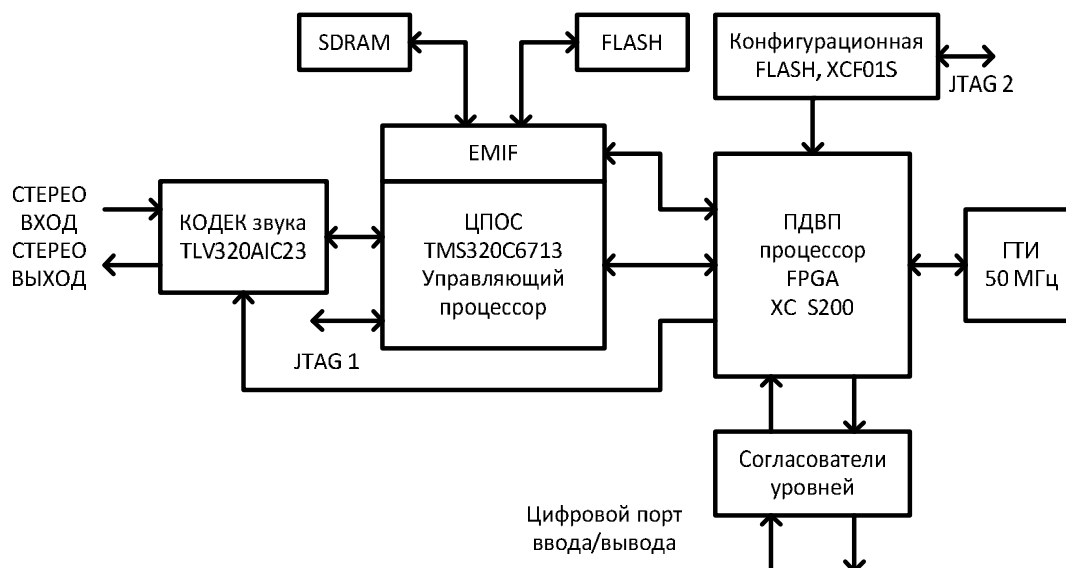


Рис. 16.16 – Структурная схема вычислительной платформы

Разработка проекта на ПЛИС реализована с помощью САПР Xilinx ISE. В данном случае конфигурационная последовательность для ПЛИС генерируется с помощью утилиты iMPACT пакета и загружается через интерфейс JTAG в конфигурационную Flash память платформы. Каждый раз после включения питания ПЛИС считывает эту последовательность из Flash памяти. Архитектурное решение вычислительной платформы позволяет получить следующие особенности: мощное вычислительное ядро с плавающей точкой для обработки данных с производительностью для 2400/1800 MIPS/MFLOPS; возможность распараллеливания задач обработки в двухпроцессорной системе; реконфигурируемость платформы путем изменения конфигурационной последовательности ПЛИС и программ ЦПОС; возможность встраивания платформы в разнообразное оборудование за счет гибкой настройки внешнего интерфейса с помощью ПЛИС; использование стандартных средств конфигурирования и программирования.

16.4.4 Технические характеристики масштабируемого компрессора звуковых сигналов

Масштабируемый компрессор звука обрабатывает кадры входного сигнала длительностью 23 мс (1024 отсчета при частоте дискретизации 44.1 кГц) на основе перцептуально адаптированного ПДВП. Перцептуально значимые вейвлет коэффициенты квантуются нелинейным квантователем с последующим кодированием без потерь при помощи кодовых книг Хаффмана. Скорость передачи для масштабируемого компрессора звука оценивается в интервале от 40 кбит/с для тональных, ритмичных звуковых сигналов до 75 кбит/с для шумовых сигналов. При этом, общая оценка соотношения шум/порог маскирования NMR_{total} меняется в пределах от

–4,35 дБ до –8,95 дБ, т.е. граница порога восприятия минимум в два раза ниже порога маскирования, при этом максимальный процент вносимых искажений в кадр реконструированного сигнала находится на уровне $MF_{max} \cong 58\%$, а ширина частотной полосы SB остается достаточной для комфортного восприятия звука и равна $SB \cong 17$ кГц.

Работа масштабируемого компрессора звука в реальном масштабе времени, на предложенной вычислительной платформе, обуславливает достаточно жесткую синхронизацию работы управляющего и ПДВП процессоров. Требуемый аппаратный ресурс для реализации ПДВП процессора на ПЛИС приведен в табл. 16.2. Перестройка ПДВП процессора осуществляется на основе вектора реконфигурации $r_{i,n}$ согласно ДТА-подхода. При этом, для обрабатываемых фреймов сигнала на ступенях ПДВП процессора, управляющий процессор осуществляет выполнение процедур расчета порогов маскирования (1440/1080 MIPS/MFLOPS), оценки перцептуальной энтропии (240/180 MIPS/MFLOPS), информативности структуры дерева ПДВП (120/90 MIPS/MFLOPS), квантование/восстановление и кодирование/декодирование (240/180 MIPS/MFLOPS).

Табл. 16.2 – Требуемые аппаратные ресурсы для реализации ПДВП процессора на ПЛИС

Ресурсы	Использовано, шт.	Израсходовано, %
Число LUT с 4-мя входами	31356	76
Число Flip Flops	3037	7
RAMB16	16	40
MULT18x18s	40	100

16.5 Список использованных источников

- Spanias A., Painter T., Atti V., Audio signal processing and coding, Hoboken, NJ: John Wiley & Sons, Inc., 2007б, 464 p.
- Bosi M., Goldberg R., Introduction to digital audio coding and standards, Springer, Inc., 2003, 434 p.
- Ковалгин Ю.А., Вологдин Э.И., Цифровое кодирование звуковых сигналов. Санкт-Петербург, Корона принт, 2004, 240 с.
- Jayant N., Noll P., Digital coding of waveforms: principles and applications to speech and video, Englewood Cliffs, NJ: Prentice-Hall, 1984, 964 p.
- Reconfigurable computing: the theory and practice of FPGA-based computing, Ed. by S.Hauck, A. Dehon, Elsevier, 2008, 908 p.
- Johnston J.D., Transform coding of audio signals using perceptual noise criteria, IEEE Trans. on Select. Areas Commun, Feb. 1988, Vol. 6, pp. 314-323.
- Wickerhauser M.V., Adaptive Wavelet Analysis from Theory to Software, A.K. Peters Ltd., Massachusetts, 1994, 486 p.

-
8. Vetterli M., Kovacevic J., Wavelets and subband coding, Englewood Cliffs, NJ, Prentice-Hall, 1995. 488 p.
 9. Petrovsky Al., Petrovsky A., Dynamic algorithm transforms for reconfigurable real-time audio coding processor, The proc. of the Intern. Conference on Parallel Computing in Electrical Engineering, PARELEC'02, Warsaw, Poland, Sep. 22-25, 2002, IEEE Press, Los Alamitos, California, 2002, pp.422-424.
 10. Petrovsky Al., Krahe D., Petrovsky A.A., Real-time wavelet packet-based low bit rate audio coding on a dynamic reconfiguration system, AES, Convention paper 5778, 114th Convention, 2003 March 22-25, Amsterdam, The Netherlands, 22 pages.
 11. Pasero E., Montuori A., Real-time performance measures of low delay perceptual audio coding, Journal of Electrical engineering, Vol. 56, No. 3-4, 2005, pp. 100-105.
 12. Montuori A., Pasero E., RTDX links for real-time performance measures of perceptual audio coding with TMS320C6713, In the proc. of the European DSP education and research symposium, Munich, April 2006, pp. 101-108.
 13. Caini C., Vanelli-Caralli A., Optimum bit allocation in subband coding with nonideal reconstruction filters, IEEE signal processing letters, Vol. 8, No. 6, June 2001, pp. 157-159.
 14. Петровский Ал.А., Построение психоакустической модели в области вейвлет коэффициентов для перцептуальной обработки звуковых и речевых сигналов. – Речевые технологии. - № 4, 2008, Москва, с. 61-71.
 15. Petrovsky Al.A., Petrovsky A.A., Reconfigurable processing: the solution to audio coding in real-time, The proc. of the 9th International conference “Mixed design of integrated circuits and systems”, MIXDES'2002, 20-22 June 2002, Wroclaw, Poland, pp. 629-634.
 16. Родионов М.М., Петровский Ал.А., Двухканальный банк фильтров на основе лестничных структур и арифметики с фиксированной запятой переменного формата, Российский научно-технический журнал "Цифровая обработка сигналов", № 2, 2010, с. 12-18.
 17. Родионов М.М., Петровский Ал.А., Поточный процессор вейвлет преобразования на базе лестничных структур с динамически реконфигурируемой архитектурой, Труды 12-й международной научной конференции “Цифровая обработка сигналов и её применение” (DSPA'2010), Том 2, Москва, Россия, 2010, с. 247-251.
 18. Vary P., Martin R., Digital Speech Transmission: Enhancement, Coding and Error Concealment, Wiley, 2006.
 19. Martin R., Petrovsky Al., Lotter T., Planar Superdirective Microphone Arrays for Speech Acquisition in the Car, 7th European conference on speech communication and technology (EUROSPEECH'2001), vol. 4, 3-7 Sept., Aalborg, Denmark, 2001, pp. 2623-2626.
 20. Петровский Ал., Перцептуальный кодер звука на базе вейвлет преобразования с динамической трансформацией частотно-временного плана, Российский научно-технический журнал "Цифровая обработка сигналов", № 4, 2009, с. 48-58.

17 ПДВП - ПРОЦЕССОР НА ЛЕСТНИЧНЫХ СТРУКТУРАХ

Родионов М.М., Петровский Ал.А.

АННОТАЦИЯ

В данной главе рассматриваются вопросы реализации алгоритма ПДВП на основе лестничных структур (lifting scheme) при использовании арифметики с фиксированной запятой. В ходе изложения материала освещаются такие вопросы как математический аппарат, позволяющий осуществить переход от алгоритма на основе КИХ фильтрации к алгоритму на базе лестничных структур, а также вопросы реализации на арифметике с фиксированной запятой. В качестве практического результата в завершении главы предлагается архитектура динамически реконфигурируемого поточного процессора ПДВП.

Ключевые слова - лестничные структуры, пакетное дискретное вейвлет преобразование, поточные процессоры, факторизация полифазной матрицы, цифровой банк фильтров.

СОДЕРЖАНИЕ

17.1 Введение.....	374
17.2 Алгоритм ПДВП.....	375
17.3 Полифазная реализация банка фильтров	376
17.4 Факторизация полифазной матрицы. Переход к лестничной структуре банка	378
17.5 Двухканальный банк фильтров анализа и синтеза на лестничных структурах	381
17.6 Вопросы реализации банка фильтров на арифметике с фиксированной запятой	383
17.7 Поточная архитектура процессора ПДВП.....	389
17.8 Операционный блок процессора	392
17.9 Список использованных источников.....	395

17.1 Введение

Мультимедиа информация с каждым днем все больше внедряется в различные сферы жизни человека. Массовое распространение приобретают цифровое телевидение и радио, видеоконференции и др. Данные обстоятельства в свою очередь требуют от разработчиков построения эффективных цифровых систем для решения таких задач, как высококачественное сжатие, редактирование шумов, а также анализ и восстановление речи, аудио, графики в реальном масштабе времени. При этом различными группами ученых ведутся комплексные разработки, направленные как на поиск новых математических алгоритмов, так и на эффективную аппаратно-программную реализацию существующих подходов.

Большой класс алгоритмов по обработке звуковой информации и графики базируется на анализе информации в частотной области, что приводит к необходимости поиска преобразования, осуществляющего заданную частотно-временную декомпозицию сигнала. В этом смысле большое распространение получило пакетное дискретное вейвлет преобразование, обладающее целым рядом преимуществ по сравнению с более традиционными подходами, например дискретным преобразованием Фурье.

Данная глава посвящена вопросам реализации алгоритма ПДВП на базе лестничных структур. Представленный ниже материал изложен в следующей последовательности. Вначале кратко рассмотрен сам алгоритм ПДВП, его древовидная структура (раздел 17.2). Затем изложена методика перехода от двухканального банка фильтров, реализованного на базе КИХ фильтрации (основного вычислительного элемента ПДВП), к банку на лестничных структурах. В разделе 17.3 приведена полифазная реализация банка, введено понятие полифазной матрицы. В разделе 17.4 изложен математический аппарат для факторизации полифазной матрицы при помощи алгоритма Евклида, и на основе этого далее приведена структура двухканального банка на лестничных структурах (раздел 17.5). В разделе 17.6 авторами рассматриваются вопросы реализации алгоритма на арифметике с фиксированной запятой. В качестве практического результата в заключение главы (раздел 17.7) предложена архитектура динамически реконфигурируемого поточного процессора ПДВП в составе перцептуального аудио кодера. В разделе 17.8 рассмотрен операционный блок процессора, реализующий функцию двухканального банка фильтров на лестничных структурах.

17.2 Алгоритм ПДВП

Схематично алгоритм ПДВП можно представить в виде древовидной структуры (рис. 17.1). В отдельно взятом узле с номером n любого уровня J дерева входной сигнал $X(z)$ сепарируется на низкочастотную $Y_L(z)$ и высокочастотную $Y_H(z)$ компоненты при помощи пары вейвлет фильтров $\tilde{h}(z)$ и $\tilde{g}(z)$ с конечной импульсной характеристикой (КИХ), после чего каждый субполосный сигнал децимируется с понижением частоты дискретизации в два раза. В литературе обычно выходные значения фильтра НЧ называют аппроксимирующими коэффициентами, а выходные значения фильтра ВЧ – детализирующими коэффициентами. Функциональный блок, реализующий данное сепарирование входного сигнала называется двухканальным банком фильтров анализа.

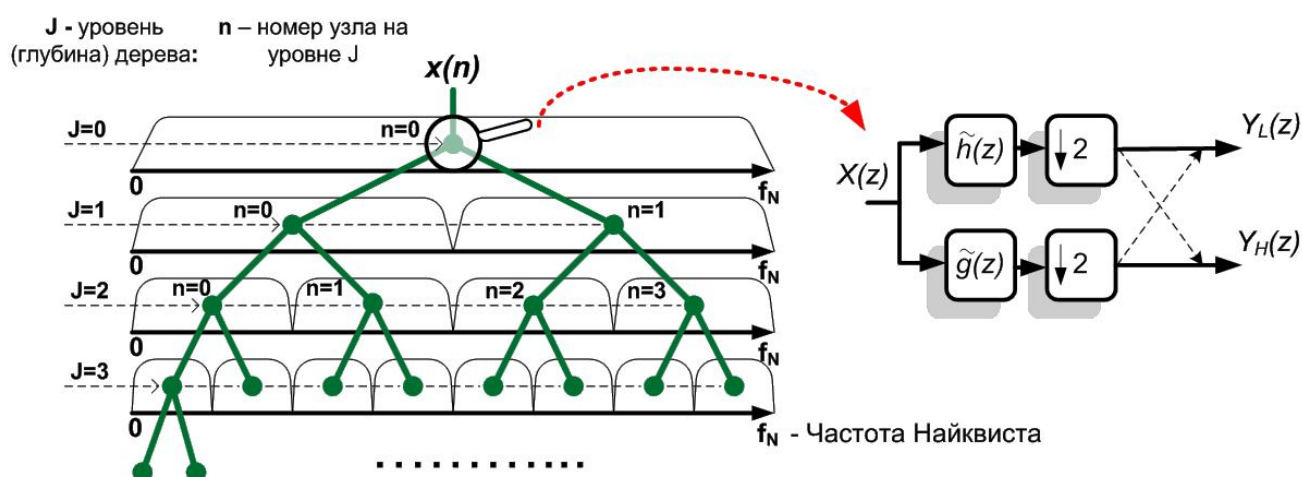


Рис. 17.1 – Древовидная структура ПДВП

Процедура обратного ПДВП реализуется на основе двухканального банка фильтров, называемого банком синтеза, который состоит из пары КИХ-фильтров $h(z)$ и $g(z)$. На рис. 17.2

представлена так называемая система «анализ-синтез», позволяющая восстановить сигнал $X(z)$ после разложение его на аппроксимирующие и детализирующие коэффициенты.

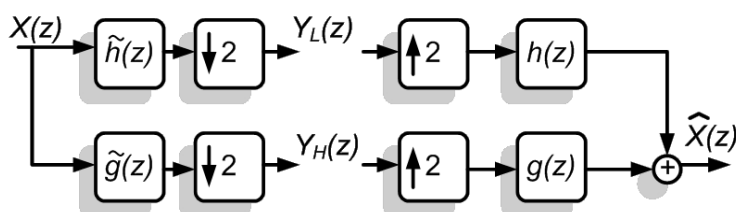


Рис. 17.2 – Система «анализ-синтез»

Таким образом, алгоритм обратного ПДВП для восстановления исходного сигнала заключается в пропускании аппроксимирующих и детализирующих коэффициентов через банк синтеза, начиная с нижних ветвей, и далее двигаясь к вершине дерева.

Для выполнения условия полного восстановления (perfect reconstruction) исходного сигнала в данной работе использованы фильтры $\tilde{h}(z)$, $\tilde{g}(z)$, $h(z)$, $g(z)$, выраженные следующей зависимостью [1]:

$$g(z) = z^{-L}h(-z^{-1}), \quad (17.1)$$

$$\tilde{h}(z) = z^{-L}h(z^{-1}), \quad (17.2)$$

$$\tilde{g}(z) = z^{-L}g(z^{-1}), \quad (17.3)$$

где $L = (N - 1)$ – порядок фильтра (N – число коэффициентов фильтра $h(z)$). В теории вейвлет преобразования $h(z)$ называют масштабирующим фильтром, частота среза которого равна половине частоты Найквиста и коэффициент усиления в полосе пропускания равен $\sqrt{2}$.

Важным преимуществом ПДВП является способность «приспосабливаться» к особенностям сигналов путем выбора соответствующей оптимальной формы дерева разложения, что в свою очередь делает алгоритм адаптивным.

17.3 Полифазная реализация банка фильтров

Вейвлет фильтры $\tilde{h}(z)$ и $\tilde{g}(z)$ в терминах z -преобразования можно определить следующими выражениями:

$$\tilde{h}(z) = \sum_{n=0}^{N-1} \tilde{h}_n z^{-n}, \quad (17.4)$$

$$\tilde{g}(z) = \sum_{n=0}^{N-1} \tilde{g}_n z^{-n}, \quad (17.5)$$

где \tilde{h}_n , \tilde{g}_n – коэффициенты фильтров $\tilde{h}(z)$ и $\tilde{g}(z)$.

Напомним, что с выходов фильтров выбираются только каждый второй отсчет, а остальные игнорируются (децимация по основанию 2). Данное обстоятельство делает возможной реализацию эффективных структур на базе КИХ-фильтров, работающих на вдвое меньшей час-

тоте относительно частоты входного сигнала. Такие структуры впервые были предложены в работе [2] и получили название полифазные.

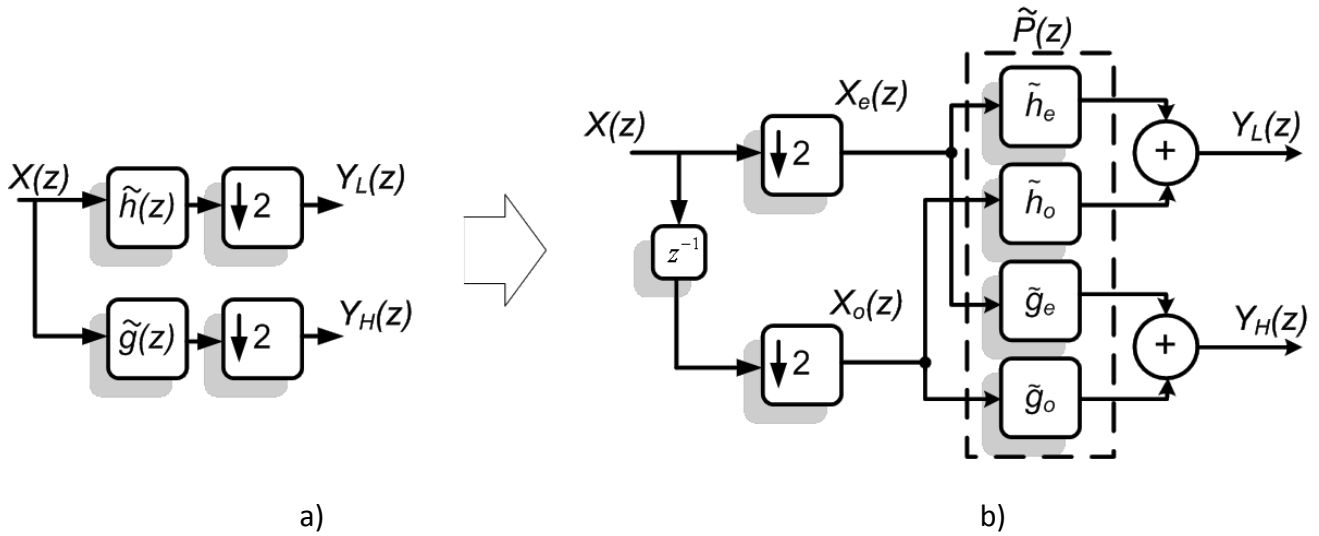


Рис. 17.3 – Структурная схема прямой (а) и полифазной (б) реализации двухканального банка фильтров

Блок-схема полифазной реализации двухканального банка фильтров изображена на рис. 17.3 б. Выходные значения $Y_L(z)$ и $Y_H(z)$ можно получить из следующего выражения:

$$\begin{bmatrix} Y_L(z) & Y_H(z) \end{bmatrix} = \begin{bmatrix} X_e(z) & z^{-1}X_o(z) \end{bmatrix} \tilde{\mathbf{P}}. \quad (17.6)$$

В данном выражении $X_e(z), X_o(z)$ – представления в z -области двух последовательно-стей, состоящих из четных и нечетных отсчетов входного сигнала, а $\tilde{\mathbf{P}}$ – полифазная матрица, формируемая на основе фильтров $\tilde{h}(z), \tilde{g}(z)$ следующим образом:

$$\tilde{\mathbf{P}} = \begin{bmatrix} \tilde{h}_e(z) & \tilde{g}_e(z) \\ \tilde{h}_o(z) & \tilde{g}_o(z) \end{bmatrix}, \quad (17.7)$$

где $\tilde{h}_e(z), \tilde{h}_o(z)$ и $\tilde{g}_e(z), \tilde{g}_o(z)$ – полифазные компоненты фильтров $\tilde{h}(z), \tilde{g}(z)$ соответственно. Ниже записаны выражения для представления полифазных компонент $\tilde{h}_e(z), \tilde{h}_o(z)$ с использованием коэффициентов фильтра $\tilde{h}(z)$ (формула (17.4)).

$$\tilde{h}_e(z) = \sum_{n=0}^{N/2-1} \tilde{h}_{2n}z^{-n}, \quad (17.8)$$

$$\tilde{h}_o(z) = \sum_{n=0}^{N/2-1} \tilde{h}_{2n+1}z^{-n}. \quad (17.9)$$

Тогда выражение, связывающее $\tilde{h}_e(z), \tilde{h}_o(z)$ и $\tilde{h}(z)$, принимает вид:

$$\tilde{h}(z) = \tilde{h}_e(z^2) + z^{-1}\tilde{h}_o(z^2). \quad (17.10)$$

Аналогичным образом записываются выражения для полифазных компонент $\tilde{g}_e(z), \tilde{g}_o(z)$.

Таким образом, процесс вычисления аппроксимирующих и детализирующих коэффициентов от входного сигнала $X(z)$ заключается в умножении вектора-строки, состоящего из элементов $X_e(z)$, $z^{-1}X_o(z)$, слева на полифазную матрицу $\tilde{\mathbf{P}}$.

17.4 Факторизация полифазной матрицы. Переход к лестничной структуре банка

Полифазная реализация не дает преимуществ в вычислительной сложности, так как число операций умножения и сложения в ней остается тем же, что и в прямой реализации.

С целью уменьшения вычислительной сложности в работах [3,4] была предложена и обоснована методика перехода от полифазной реализации к схеме двухканального банка фильтров на основе лестничных структур (в зарубежной литературе *lifting scheme*). Эффективность предложенного подхода послужила толчком для развития и широкого использования лестничных структур для реализации алгоритмов ДВП в системах мультимедиа.

Для перехода к предложенной архитектуре осуществляется факторизация полифазной матрицы $\tilde{\mathbf{P}}$, т.е. разложение на более простые треугольные матрицы, называемыми элементарными шагами.

$$\mathbf{S} = \begin{bmatrix} 1 & s(z) \\ 0 & 1 \end{bmatrix}, \quad (17.11)$$

$$\mathbf{T} = \begin{bmatrix} 1 & 0 \\ t(z) & 1 \end{bmatrix}. \quad (17.12)$$

При этом матрица \mathbf{S} называется прямым шагом, а матрица \mathbf{T} дуальным шагом [4].

Прежде чем подробно рассмотреть алгоритм факторизации, введем некоторые математические определения и выражения.

Полином Лорана. Полиномом Лорана называют полином вида [5]:

$$a(z) = \sum_{k=m}^n a_k z^{-k}, \quad (17.13)$$

где m, n – целые числа, $m \leq n$, a_k – вещественные коэффициенты, если a_m и a_n отличны от нуля, то число $|h| = n - m$ называют степенью полинома Лорана $h(z)$.

Таким образом, можно утверждать, что передаточные функции КИХ-фильтров $\tilde{h}(z)$ и $\tilde{g}(z)$ являются полиномами Лорана.

Теперь введем понятие комплементарности двух полиномов Лорана. Два полинома Лорана $h(z)$ и $g(z)$ являются комплементарными, если определитель их полифазной матрицы (17.7) тождественно равен 1.

Отметим следующее свойство, касающееся комплементарных полиномов. Если пары полиномов Лорана $h(z), g(z)$ и $h(z), g^0(z)$ – комплементарные, а \mathbf{P} и \mathbf{P}^0 их полифазные матрицы, то справедливо следующее равенство:

$$\mathbf{P} = \mathbf{P}^0 \begin{bmatrix} 1 & s(z) \\ 0 & 1 \end{bmatrix}, \quad (17.14)$$

где

$$s(z) = g_e(z)g_o^0(z) - g_o(z)g_e^0(z). \quad (17.15)$$

На основании вышеизложенного перейдем к рассмотрению непосредственно самого метода для факторизации полифазной матрицы (17.7). В основе данного подхода лежит алгоритм Евклида [5], позволяющий разложить столбец из элементов $\tilde{h}_e(z)$ и $\tilde{h}_o(z)$ на произведение матриц следующего вида:

$$\begin{bmatrix} \tilde{h}_e(z) \\ \tilde{h}_o(z) \end{bmatrix} = \prod_{i=1}^N \begin{bmatrix} q_i(z) & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} K \\ 0 \end{bmatrix}, \quad (17.16)$$

где N – число шагов в алгоритме Евклида, $q_i(z)$ – последовательные неполные частные (полиномы Лорана), K – вещественное число, являющееся наименьшим общим делителем полиномов $\tilde{h}_e(z)$ и $\tilde{h}_o(z)$.

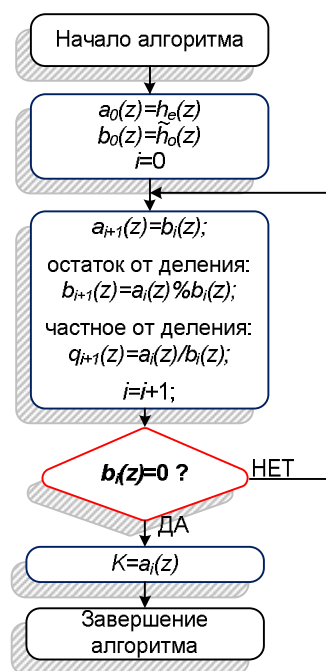


Рис. 17.4 – Алгоритм Евклида

Алгоритм Евклида может быть записан в виде следующей последовательности действий (рис. 17.4):

- 1) присвоение значений полиномов $\tilde{h}_e(z)$ и $\tilde{h}_o(z)$ переменным $a_0(z)$ и $b_0(z)$, и обнуление индекса i .
- 2) присвоение переменной $a_{i+1}(z)$ значения $b_i(z)$ (на первом шаге алгоритма $a_1(z) = b_0(z) = \tilde{h}_o(z)$).
- 3) нахождение частного от деления полиномов $a_i(z)$ на $b_i(z)$: $q_{i+1}(z) = a_i(z)/b_i(z)$.
- 4) присвоение переменной $b_{i+1}(z)$ остатка от деления полиномов $a_i(z)$ на $b_i(z)$.
- 5) увеличение индекса итераций на 1: $i = i + 1$.
- 6) проверка на равенство нулю остатка от деления полиномов на текущей итерации: $(b_i(z) == 0)?$. Если нет, то повтор выполнения пунктов 3-6 алгоритма, в противном случае завершение алгоритма.

В результате получаем значения полиномов $q_i(z)$ и значение коэффициента $K = a_i(z)$ для формулы (17.16).

Введем полиномы $\tilde{g}_e^0(z)$ и $\tilde{g}_o^0(z)$ с помощью формулы:

$$\begin{bmatrix} \tilde{g}_e^0(z) \\ \tilde{g}_o^0(z) \end{bmatrix} = \prod_{i=1}^N \begin{bmatrix} q_i(z) & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ K^{-1} \end{bmatrix}. \quad (17.17)$$

При этом согласно [4] полиномы $\tilde{h}(z)$ и $\tilde{g}^0(z)$, где

$$\tilde{g}^0(z) = \tilde{g}_e^0(z^2) + z^{-1}\tilde{g}_o^0(z^2), \quad (17.18)$$

являются комплементарными, и их полифазная матрица имеет вид

$$\tilde{\mathbf{P}}^0 = \begin{bmatrix} \tilde{h}_e(z) & \tilde{g}_e^0(z) \\ \tilde{h}_o(z) & \tilde{g}_o^0(z) \end{bmatrix} = \prod_{i=1}^N \begin{bmatrix} q_i(z) & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} K & 0 \\ 0 & K^{-1} \end{bmatrix}. \quad (17.19)$$

Теперь на основании (17.14) можем осуществить переход от полифазной матрицы $\tilde{\mathbf{P}}^0$ к полифазной матрице $\tilde{\mathbf{P}}$ исходных фильтров $\tilde{h}(z)$ и $\tilde{g}(z)$

$$\tilde{\mathbf{P}} = \tilde{\mathbf{P}}^0 \begin{bmatrix} 1 & s(z) \\ 0 & 1 \end{bmatrix}, \quad (17.20)$$

где $s(z)$ – полином Лорана, рассчитанный из формулы (17.15). Таким образом, получаем следующее выражение для $\tilde{\mathbf{P}}$:

$$\tilde{\mathbf{P}} = \prod_{i=1}^N \begin{bmatrix} q_i(z) & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} K & 0 \\ 0 & K^{-1} \end{bmatrix} \begin{bmatrix} 1 & s(z) \\ 0 & 1 \end{bmatrix}. \quad (17.21)$$

Для представления выражения (17.21) в другом виде используются следующие свойства матриц:

$$\begin{bmatrix} q(z) & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & q(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad (17.22)$$

$$\begin{bmatrix} q(z) & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ q(z) & 1 \end{bmatrix}, \quad (17.23)$$

$$\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad (17.24)$$

$$\begin{bmatrix} K & 0 \\ 0 & K^{-1} \end{bmatrix} \begin{bmatrix} 1 & s(z) \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & K^2 s(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} K & 0 \\ 0 & K^{-1} \end{bmatrix}. \quad (17.25)$$

В итоге получаем факторизацию полифазной матрицы следующего вида:

$$\tilde{\mathbf{P}} = \prod_{i=1}^{N/2} \left(\begin{bmatrix} 1 & s_i(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ t_i(z) & 1 \end{bmatrix} \begin{bmatrix} 1 & K^2 s(z) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} K & 0 \\ 0 & K^{-1} \end{bmatrix} \right), \quad (17.26)$$

где $s_i(z) = q_{2i-1}(z)$, $t_i(z) = q_{2i}(z)$.

Теперь рассмотрим использование описанного алгоритма факторизации применительно к двухканальному банку фильтров в ПДВП.

17.5 Двухканальный банк фильтров анализа и синтеза на лестничных структурах

Для выполнения факторизации данной матрицы и последующего моделирования работы двухканального банка фильтра на лестничных структурах была написана программа в пакете MATLAB [6].

В итоге общий вид полиномов $s_i(z)$ и $t_i(z)$ свелся к следующему выражению:

$$q_i(z) = (b_0^i + b_1^i z^{-1})z^i, i = 1 \dots \left(\frac{N}{2} + 1\right), \quad l \in Z. \quad (17.27)$$

На основании представленного выражения двухканальный банк фильтров структурно можно изобразить в виде блок схемы, изображенной на рис. 17.5.

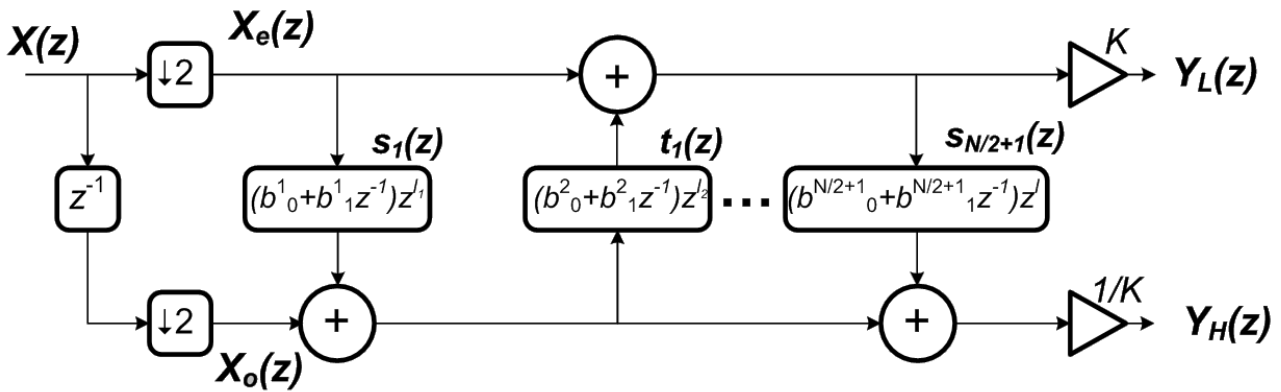


Рис. 17.5 – Двухканальный банк анализа на лестничных структурах

Очевидно, что для восстановления сигнала необходим банк синтеза со структурой, изображенной на рис. 17.6.

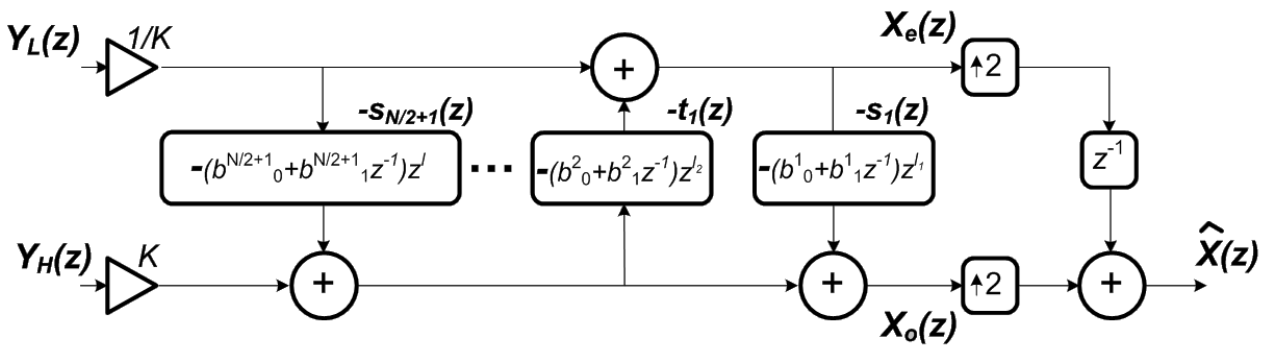


Рис. 17.6 – Двухканальный банк синтеза на лестничных структурах

В матричном виде выражение для получения восстановленного сигнала принимает вид:

$$\begin{aligned} [\hat{X}_e(z) \quad \hat{X}_o(z)] &= \\ &= [Y_L(z) \quad Y_H(z)] \begin{bmatrix} K^{-1} & 0 \\ 0 & K \end{bmatrix} \begin{bmatrix} 1 & -K^2 s(z) \\ 0 & 1 \end{bmatrix} \prod_{i=\frac{N}{2}}^1 \left(\begin{bmatrix} 1 & 0 \\ -t_i(z) & 1 \end{bmatrix} \begin{bmatrix} 1 & -s_i(z) \\ 0 & 1 \end{bmatrix} \right). \end{aligned} \quad (17.28)$$

В данном выражении $\hat{X}_e(z)$, $\hat{X}_o(z)$ – представления в z -области последовательностей, состоящих из четных и нечетных отсчетов восстановленного сигнала.

Для лучшего понимания изложенного материала рассмотрим пример с рассчитанными параметрами лестничных структур (коэффициентов b_0^i , b_1^i , K , K^{-1}) для фильтров Добеши с числом коэффициентов равным 12 (одиннадцатый порядок). Ниже в табл. 17.1 приведены коэффициенты исходных фильтров. Далее в табл. 17.2 – коэффициенты b_0^i , b_1^i , K , K^{-1} и значение степени l для лестничных структур.

Табл. 17.1 – Коэффициенты фильтров $\tilde{h}(z)$ и $\tilde{g}(z)$ семейства Добеши 11 порядка

$\tilde{h}(z)$	Значение коэффициента	$\tilde{g}(z)$	Значение коэффициента
\tilde{h}_0	-0,00107730108530848	\tilde{g}_0	-0.111540743350110
\tilde{h}_1	+0,00477725751094552	\tilde{g}_1	+0.494623890398454
\tilde{h}_2	+0,000553842201161498	\tilde{g}_2	-0.751133908021096
\tilde{h}_3	-0,0315820393174861	\tilde{g}_3	+0.315250351709198
\tilde{h}_4	+0,0275228655303056	\tilde{g}_4	+0.226264693965440
\tilde{h}_5	+0,0975016055873230	\tilde{g}_5	-0.129766867567262
\tilde{h}_6	-0.129766867567262	\tilde{g}_6	-0.097501605587323
\tilde{h}_7	-0.226264693965440	\tilde{g}_7	+0.027522865530306
\tilde{h}_8	0.315250351709198	\tilde{g}_8	+0.031582039317486
\tilde{h}_9	0.751133908021096	\tilde{g}_9	+0,000553842201161498
\tilde{h}_{10}	0.494623890398454	\tilde{g}_{10}	-0.004777257510946
\tilde{h}_{11}	0.111540743350110	\tilde{g}_{11}	+0.111540743350110

Табл. 17.2 – Параметры лестничной структуры при исходных фильтрах $\tilde{h}(z)$ и $\tilde{g}(z)$ семейства Добеши 11 порядка

	$s_1(z)$	$t_1(z)$	$s_2(z)$	$t_2(z)$
b_0^i	-4,43446829869067	-0,0633131925460941	9.970015612798715	-0.023663493662427
b_1^i	0	0,214593450003008	-4.493113168870837	0.057413936848331
l	0	1	-1	3

	$s_3(z)$	$t_3(z)$	$s_4(z)$	K
b_0^i	2.356497021976010	-0,0009911655305195	0.094106674041988	3.121464721105518
b_1^i	-0.67878434614780	0.007183563115834		K^{-1}
l	-3	5	-5	0.320362422563544

На рис. 17.7 приведена структурная схема для данного примера. Так как мы работаем с казуальными системами, т. е. в определенный момент времени значение выходного отсчета зависит только от текущего и предыдущих входных значений (в случае нерекурсивных систем), то между элементарными шагами лестничной структуры введены элементы задержки для «компенсации» положительных степеней от переменной z .

В рассмотренном примере число операций умножения составляет 14, а число операций сложения – 12. В свою очередь при прямой (либо полифазной) реализации значения данных параметров равны 24 и 22 соответственно. В ходе анализа полученных результатов была

выведена следующая зависимость числа умножений N_{mul} и числа сложений N_{add} от количества коэффициентов N фильтра $\tilde{h}(z)$:

$$N_{mul} = N + 2, \quad (17.29)$$

$$N_{add} = N. \quad (17.30)$$

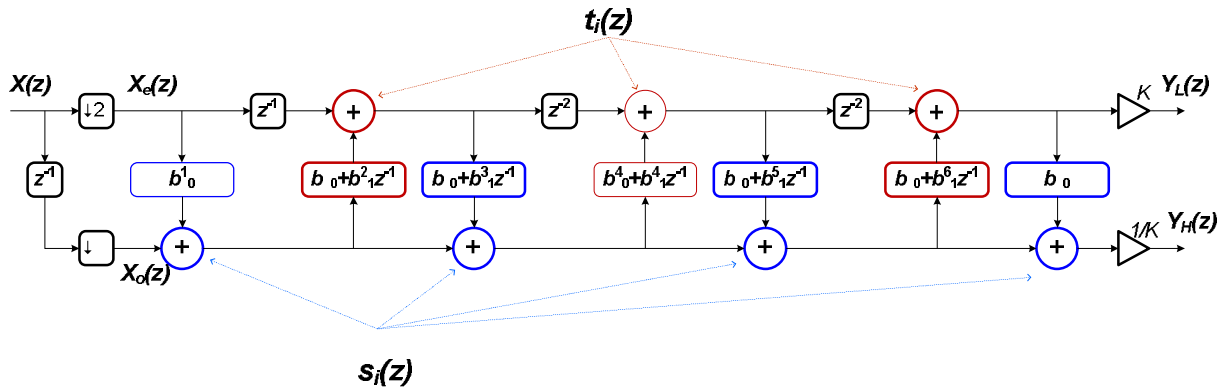


Рис. 17.7 – Блок-схема двухканального банка фильтров анализа на лестничных структурах для исходных фильтров Добеши 11-го порядка

В случае прямой реализации, формулы, выражающие ту же зависимость, имеют следующий вид:

$$N_{mul} = 2N, \quad (17.31)$$

$$N_{add} = 2(N - 1). \quad (17.32)$$

Таким образом, вычислительная сложность банка фильтров на лестничных структурах почти в 2 раза меньше по сравнению с прямой реализацией.

17.6 Вопросы реализации банка фильтров на арифметике с фиксированной запятой

При проектировании портативных цифровых систем, работающих в реальном масштабе времени, особое внимание уделяется таким показателям, как скорость вычислений, аппаратные затраты, сложность алгоритмов выполнения арифметических операций. По этой причине большинство устройств такого класса выполняются на основе арифметики с фиксированной запятой, которая позволяет получить хорошие результаты по перечисленным показателям. Ниже рассмотрена методика для реализации алгоритма ПДВП на лестничных структурах с использованием данной арифметики. При этом значения входного сигнала представляются в виде дробных чисел в дополнительном коде.

Проведем анализ математической модели двухканального банка фильтров на лестничных структурах.

Коэффициенты b_0^i и b_1^i , полученные в результате факторизации полифазной матрицы, могут выходить за пределы диапазона $[-1, 1)$. Это в свою очередь приводит к снижению вход-

ного динамического диапазона, если использовать арифметику с фиксированной запятой для кодирования только дробных чисел. Поэтому в данной работе было принято решение использовать так называемую арифметику с переменным количеством бит, отводимых под целую и дробную части числа [7,8].

В соответствии с данной методикой любое число, представленное в формате с фиксированной запятой в дополнительном коде, задается двумя параметрами: длина слова wl (word length), и число бит, отводимых под целую часть iwl (integer word length). Оставшееся число бит fwl (fractional word length) отводится под кодирование дробной части (рис. 17.8).

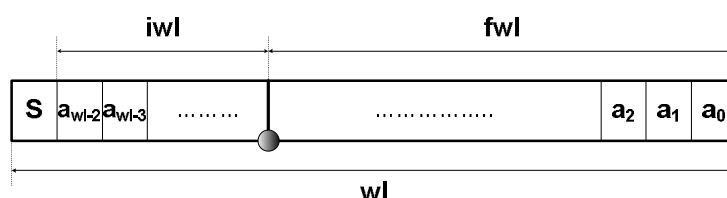


Рис. 17.8 – Представление числа в формате с фиксированной запятой

Таким образом, любое число, заданное в таком формате, может быть записано в следующей форме:

$$a = ma \cdot 2^{iwl}, \quad (17.33)$$

где ma представляется, как

$$ma = (-1)^s + \sum_{i=0}^{wl-2} a_i \cdot 2^{i-wl+1}. \quad (17.34)$$

В данном выражении a_i – это значения соответствующих битов числа, равные 0 либо 1, s – знаковый бит.

Рассмотрим особенности операций сложения и умножения.

Пусть на определенном участке алгоритма необходимо произвести сложение двух входных переменных, заданных в следующих форматах: $a (wl, iwla)$, $b (wl, iwlb)$, где wl – длина слова, а $iwla, iwlb$ – число бит, отведенных под целую часть, при этом пусть $iwlb \geq iwla$, тогда результат суммы с учетом (17.33) можно записать как:

$$c = a + b = mc \cdot 2^{iwlc} = (ma \cdot 2^{iwla-iwlb} + mb) \cdot 2^{iwlb}. \quad (17.35)$$

В данной записи умножение на $2^{iwla-iwlb}$ соответствует арифметическому сдвигу вправо для «выравнивания» битов с одинаковыми весами в операндах a и b (рис. 17.9).

Для большинства цифровых систем длина слова wl является фиксированной, поэтому младшие $iwlb - iwla$ бит числа a в результате арифметического сдвига отбрасываются и не участвуют в операции сложения.

Теперь рассмотрим случай, когда необходимо для тех же операндов a и b выполнить операцию умножения (Рис. 17.10).

В соответствии с (17.33) результат умножения запишем как:

$$c = mc \cdot 2^{iwlc} = ma \cdot mb \cdot 2^{iwla+iwlb}. \quad (17.36)$$

На основании (17.34) можно отметить, что значения $ma, mb \in [-1,1)$, поэтому операцию $ma \cdot mb$ можно выполнять по правилам умножения дробных чисел, представленных в дополнительном коде. Тогда результат умножения будет иметь длину слова равную $2 \cdot wl$, при этом первый значащий бит сместится на одну позицию вправо. Таким образом, корректное значение mc для выбранной длины слова wl будет расположено в разрядах, начиная с $wl - 1$ и по $2wl - 2$.

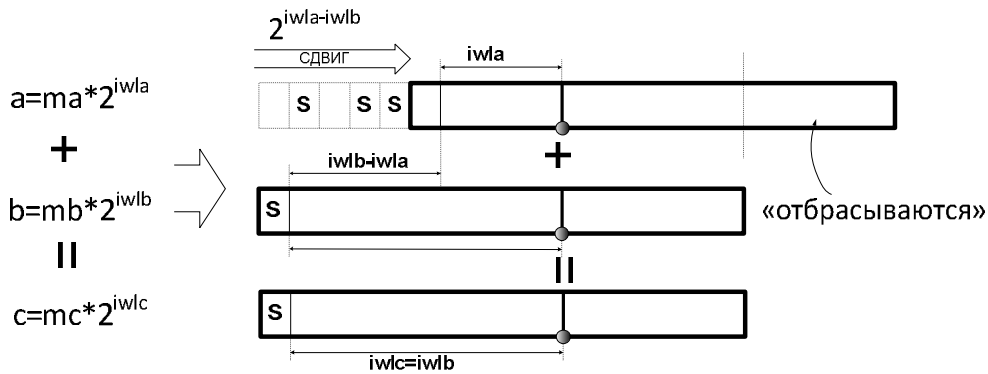


Рис. 17.9 – Сложение чисел с различным числом бит, отведенных для кодирования целой части

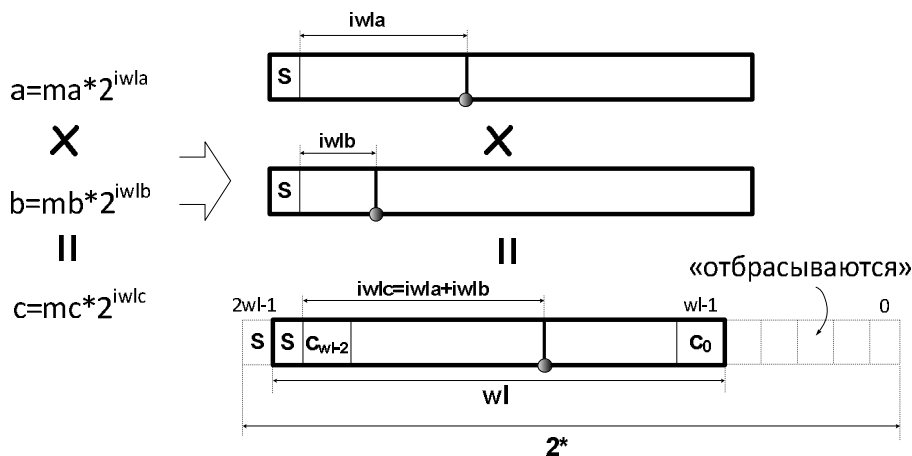


Рис. 17.10 – Умножение чисел с различным числом бит, отведенных для кодирования целой части

Основываясь на вышеизложенный материал, перейдем к рассмотрению самой реализации двухканального банка фильтров на арифметике с фиксированной запятой. Основной шаг лестничной структуры банка может быть представлен в виде блок-схемы, изображенной на рис. 17.11. На вход рассматриваемой схемы поступают данные $X_e(n)$ и $X_o(n)$ с одинаковой разрядностью, но с различным числом бит отведенных под целую часть числа. Возможен случай, когда целая часть будет отсутствовать в формате числа, и, более того, старший значащий бит числа может иметь вес 2^{-k} , где k – целое положительное. Тогда более корректной будет следующая запись:

$$X_e = m_{X_e} \cdot 2^{exp_{X_e}} \quad (17.37)$$

где exp_{X_e} – это порядок масштабирующего множителя $2^{exp_{X_e}}$, который необходим для правильной интерпретации значения числа X_e (либо X_o). С точки зрения расчетных формул (17.33) – (17.36) iwl и exp_A представляют собой одно и то же, но по смыслу это разные величины.

Таким образом, далее в работе для задания в алгоритме произвольной переменной (константы) X с длиной слова wl будут использоваться два параметра: $m_X \in [-1; 1]$ – значение переменной (константы) без учета масштабирующего множителя; $exp_X \in Z$ – порядок масштабирующего множителя.

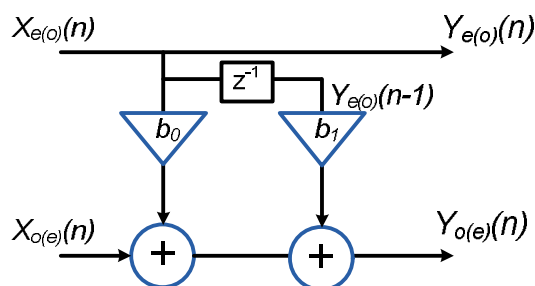


Рис. 17.11 – Блок-схема основного шага в двухканальном банке фильтров на лестничных структурах

Для выполнения арифметических вычислений необходимо выполнить квантование (перевод из формата с плавающей запятой в формат с фиксированной запятой) коэффициентов b_0 и b_1 . В данной работе было принято решение осуществить квантование на основе предложенного выражения:

$$b = m_b \cdot 2^{exp_b}, |m_b| \in [0.5; 1], exp_b \in Z. \quad (17.38)$$

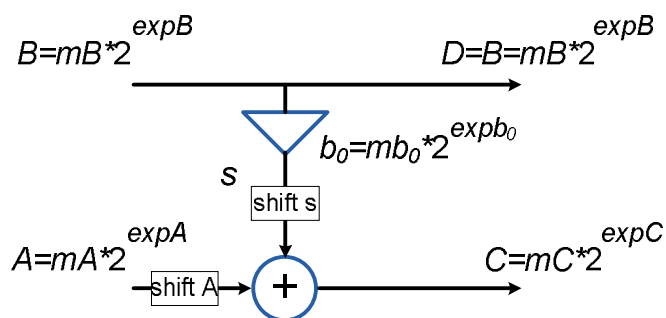


Рис. 17.12 – Основной функциональный блок банка фильтров

В сущности, данное выражение подразумевает, что при переводе из формата с плавающей запятой в формат с фиксированной запятой значению m_b присваивается мантисса исходного коэффициента, заданного в формате с плавающей запятой, а параметру exp_b – порядок этого значения. Данный способ позволяет сохранить наибольшую точность при переводе числа при выбранной длине слова.

Как видно из рис. 17.11, основным функциональным блоком (рис. 17.12) в нашем случае является умножение одной входной переменной на постоянный коэффициент и последующее сложение получившегося результата со второй входной переменной.

Ниже приведены выражения для расчета промежуточных и выходных значений в данном функциональном блоке.

Расчет параметров результата умножения $s = B \cdot b_0$:

$$m_s = m_B \cdot m_{b_0}, \quad (17.39)$$

$$exp_s = exp_B + exp_{b_0}. \quad (17.40)$$

Расчет параметров выходной переменной $C = A + s$:

$$exp_C = \max(exp_A, exp_s), \quad (17.41)$$

$$m_A = m_A \cdot 2^{exp_C - exp_A} + m_s \cdot 2^{exp_C - exp_s}. \quad (17.42)$$

В итоге вся процедура расчета сводится к вычислению двух арифметических операций над дробными знаковыми числами, представленными в дополнительном коде:

- произведение входного значения m_B на константу m_{b_0} (17.39):
- сложение результата умножения m_s и второго входного значения m_A , причем операнд с меньшим значением порядка масштабирующего множителя предварительно подвергается арифметическому сдвигу вправо на $exp_C - exp_A$ (exp_s) бит.

Таким образом, для реализации рассматриваемого алгоритма на арифметике с фиксированной запятой необходимо выполнить следующие действия: квантование коэффициентов лестничных структур b_0^i, b_1^i в соответствии с формулой (17.38), расчет параметров арифметического сдвига операндов перед выполнением операций суммирования.

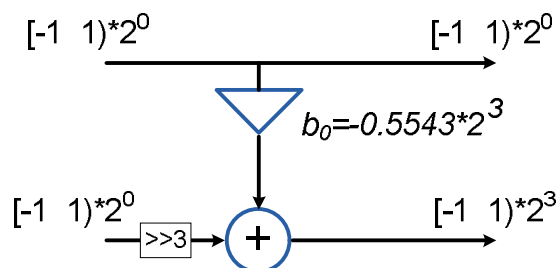


Рис. 17.13 – Реализация на арифметике с фиксированной запятой первого шага двухканального банка фильтров анализа

Рассмотрим следующий пример. Возьмем параметры первого шага лестничной структуры для фильтра Добеши 11-го порядка из табл. 17.2: $s_1(z) = b_0^1 = -4,43446829869067$. После квантования параметры коэффициента b_0^1 будут равны $m_{b_0} = -0,5543$ и $exp_{b_0} = 3$. Так как на первом шаге лестничной структуры на вход поступают последовательности $X_e(z)$ и $X_o(z)$, состоящие из четных и нечетных отсчетов входного сигнала $X(z)$, то для них $exp_{X_e} = exp_{X_o} = exp_X = 0$, а параметры $m_{X_e}, m_{X_o} \in [-1; 1)$ есть сами значения входных данных. На основе формул (17.39)–(17.42), принимая, что $A = X_o$, а $B = X_e$ получаем, что $exp_C = 3$. Поэтому перед операцией сложения необходимо произвести арифметический сдвиг вправо операнда m_{X_o} на три разряда (см. рис. 17.13).

И так, продвигаясь от первого шага лестничной структуры до последнего, можно рассчитать значения порядков масштабирующих коэффициентов для всех узлов схемы, а также определить параметры арифметических сдвигов для операндов перед операцией суммирования.

Теперь, когда осуществлено преобразование основных шагов исходного алгоритма на арифметику с фиксированной запятой, осталось рассмотреть операцию умножения на диагональную матрицу, состоящую из коэффициентов K и K^{-1} (см. формулу (17.26)). Как упоминалось ранее, исходная полифазная матрица формируется на основе фильтров $\tilde{h}(z)$ и $\tilde{g}(z)$. Причем данная пара фильтров имеет коэффициент усиления в полосе пропускания равный $\sqrt{2}$. А это значит, что и банк фильтров, реализованный на лестничных структурах, будет иметь коэффициент усиления в НЧ и ВЧ канале тоже равный $\sqrt{2}$. Поэтому для нормировки к единичному коэффициенту усиления заменим выходную матрицу коэффициентов из формулы на матрицу вида:

$$K = \begin{bmatrix} K_1 & 0 \\ 0 & K_2 \end{bmatrix} = \begin{bmatrix} \frac{K}{\sqrt{2}} & 0 \\ 0 & \frac{K^{-1}}{\sqrt{2}} \end{bmatrix}. \quad (17.43)$$

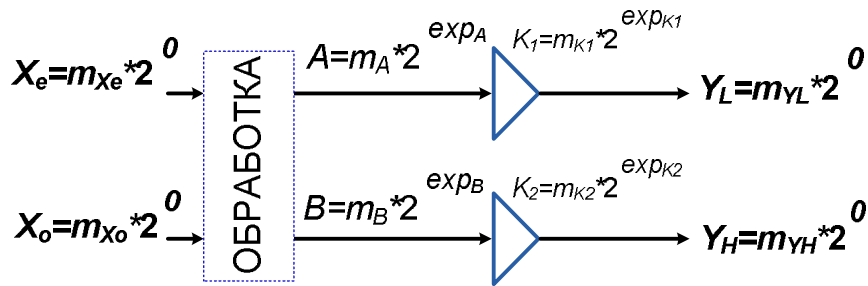


Рис. 17.14 – Схема умножения на матрицу коэффициентов K

Операцию умножения вектора-строки промежуточных данных $[A \ B]$ на матрицу K (рис. 17.14) можно представить в виде следующих выражений:

$$Y_L = m_{YL} \cdot 2^0 = m_A \cdot m_{K1} \cdot 2^{exp_A + exp_{K1}}, \quad (17.44)$$

$$Y_H = m_{YH} \cdot 2^0 = m_B \cdot m_{K2} \cdot 2^{exp_B + exp_{K2}}. \quad (17.45)$$

Порядок масштабирующих множителей для выходных значений Y_L и Y_H $exp_{YL} = exp_{YH} = 0$, так как общая схема алгоритма имеет коэффициент усиления не превышающий единицы на всем частотном диапазоне, т.е. вес старшего разряда в выходном слове равняется 2^0 . Это в свою очередь должно означать, что $exp_A + exp_{K1} = exp_B + exp_{K2} = 0$, однако в действительности выражения $exp_A + exp_{K1}$ ($exp_B + exp_{K2}$) ≥ 0 .

Таким образом, для того чтобы получать на выходе двухканального банка анализа значения с длиной слова, равной wl , и весом старшего бита 2^0 , в данной работе был использован следующий прием. В результате перемножения значений A и K_1 (B и K_2) мы получаем значения с разрядностью $2 \cdot wl$ и весом старшего бита $2^{1+exp_A+exp_{K1}}$ ($2^{1+exp_B+exp_{K2}}$). Поскольку выходные значения лежат в интервале $[-1, 1)$, то разряды с $2wl - 1$ по $2wl - 1 - exp_A - exp_{K1}$ ($2wl - 1 - exp_B - exp_{K2}$) не являются значащими, и лишь «расширяют» знаковый бит (рис. 17.15). Таким образом, выходной результат располагается в разрядах с $2wl - 2 - exp_A - exp_{K1}$

по $wl - 1 - exp_A - exp_{K1}$ для результата умножения $Y_L = A \cdot K_1$, и в разрядах с $2wl - 2 - exp_B - exp_{K2}$ по $wl - 1 - exp_B - exp_{K2}$ для результата умножения $Y_H = B \cdot K_2$.

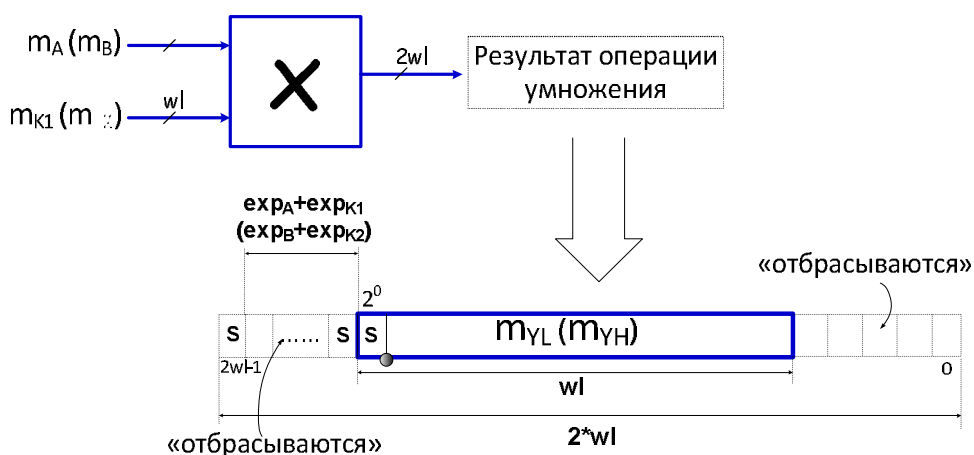


Рис. 17.15 – Формирование выходного значения Y_L (Y_H) из результата операции умножения значений A и K_1 (B и K_2)

В результате получена математическая модель двухканального банка фильтров анализа для арифметики с фиксированной запятой.

17.7 Поточная архитектура процессора ПДВП

Для эффективной реализации алгоритма ПДВП большое распространение получили поточные архитектуры [9-15]. В данной работе в качестве аппаратной реализации для алгоритма ПДВП на основе лестничных структур предлагается динамически реконфигурируемый поточный процессор на базе ПЛИС.

Представим алгоритм ПДВП в виде сигнального ориентированного графа. На данном графе входной сигнал x представляется как набор значений $x[n]$, где $n = 0, 1, 2 \dots$ – это номер входного отсчета. Расчет очередных двух значений с выхода двухканального банка фильтров производится с помощью вычислительного элемента ВЭ, на вход которого поступают два отсчета входного сигнала (либо два промежуточных отсчета, лежащих в одной частотной субполосе). Также для данного графа введены следующие обозначения:

- $J = 0..7$ – номер уровня декомпозиции;
- $w_{J,k}[n], k = 0..(2^{J+1} - 1)$ – n -ый вейвлет коэффициент k -ой субполосы, полученный на J -ом уровне декомпозиции;
- $G_{k,i}$ – идентификатор i -го ВЭ, обрабатывающего последовательность из k -ой субполосы, полученной на предыдущем уровне декомпозиции, т.е. для ВЭ на J -ом уровне индекс k может принимать значения от 0 до 2^J .

На каждом уровне декомпозиции J все ВЭ разбиваются на 2^J группы так, чтобы вычислительные элементы в одной группе обрабатывали последовательность только из одной частотной субполосы. В идентификаторе $G_{k,i}$ индекс k обозначает номер группы, а индекс i – номер ВЭ в k -ой группе. Таким образом, идентификатор $G_{k,i}$ уникален для каждого ВЭ в рамках одного уровня декомпозиции. В качестве примера на рисунке рис. 17.16 приведен сигнальный

ориентированный граф для алгоритма полного трехуровневого ПДВП (т.е. для алгоритма, описывающего структуру полного дерева, в котором присутствуют все ветви).

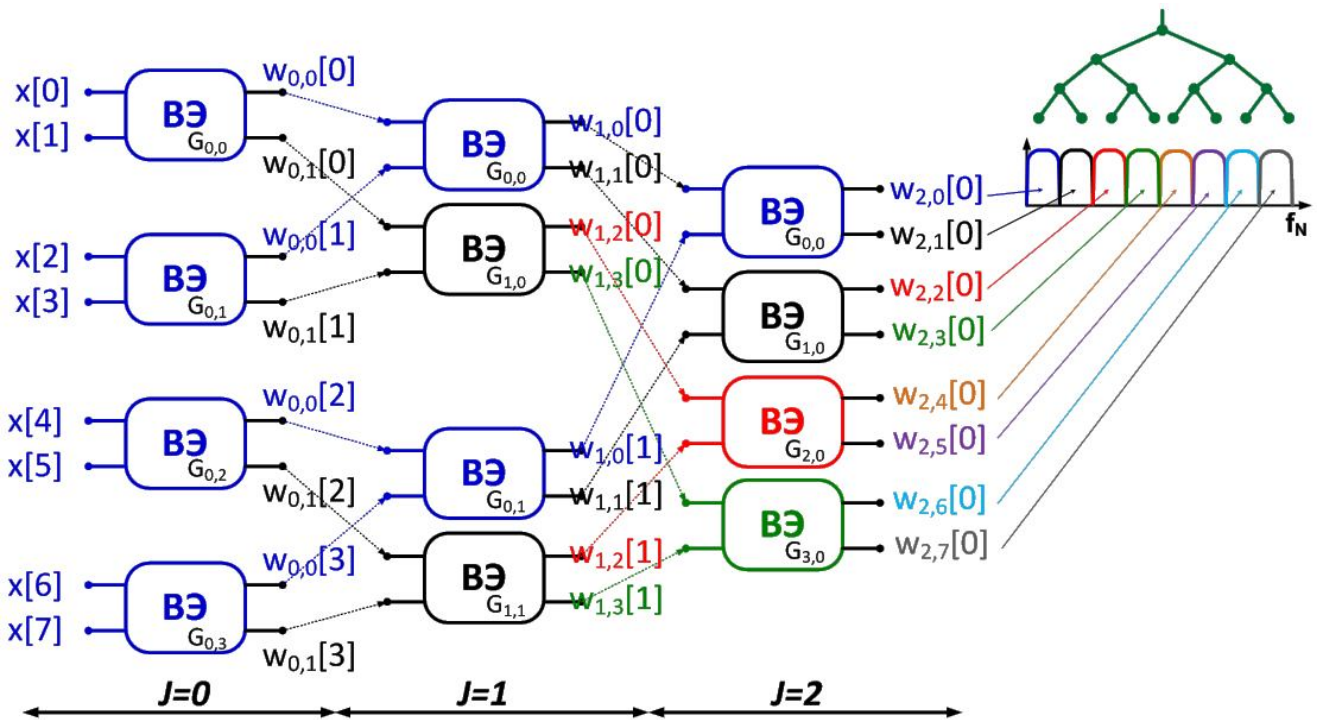


Рис. 17.16 – Сигнальный ориентированный граф для алгоритма полного трехуровневого ПДВП

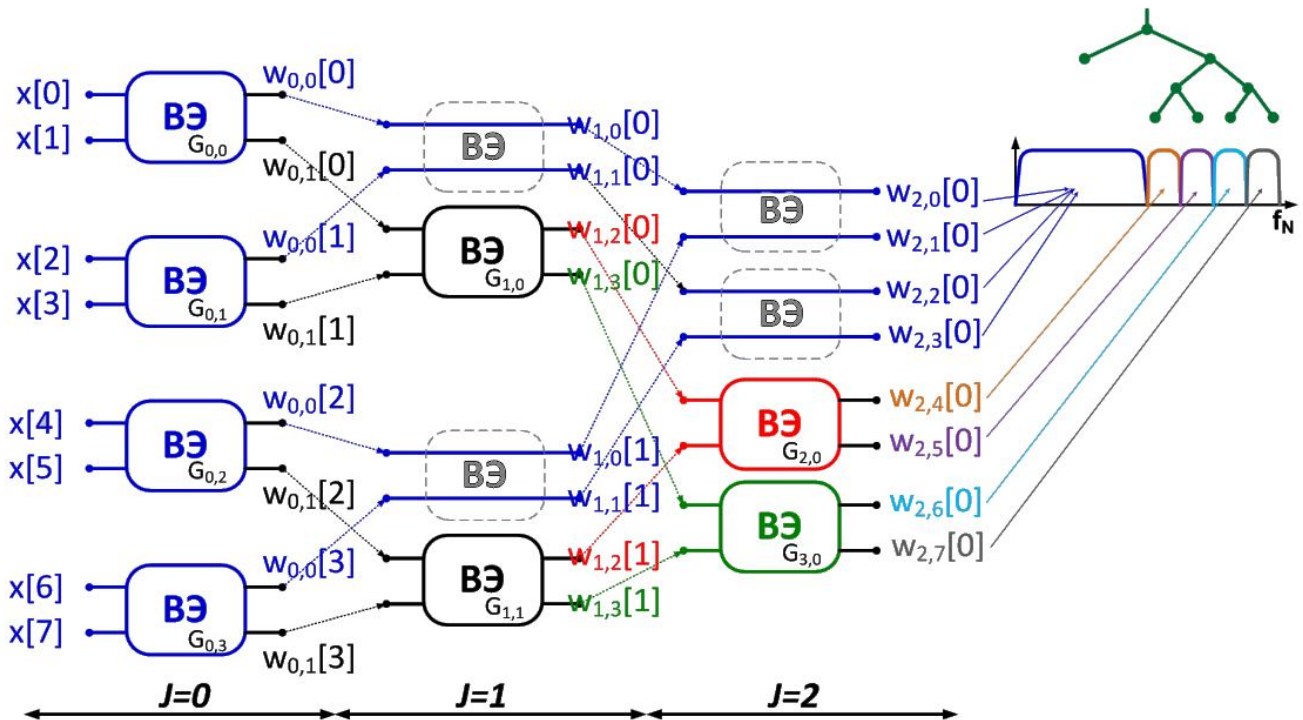


Рис. 17.17 – Сигнальный ориентированный граф алгоритма трехуровневого ПДВП для неполной структуры дерева.

С помощью представленного графа можно описывать не только алгоритм полного ПДВП, но и алгоритм с произвольной структурой дерева. Для этого необходимо в тех местах,

где отсутствуют ветви в структуре дерева, просто убрать ВЭ, а данные, которые должны были бы обрабатываться этими ВЭ, пропустить без изменения на следующий уровень графа.

На рис. 17.17 приведен пример для трехуровневого ПДВП с неполной структурой дерева.

С учетом анализа алгоритма ПДВП на основе представленного графа предлагается точная архитектура процессора ПДВП, показанная на рис. 17.18.

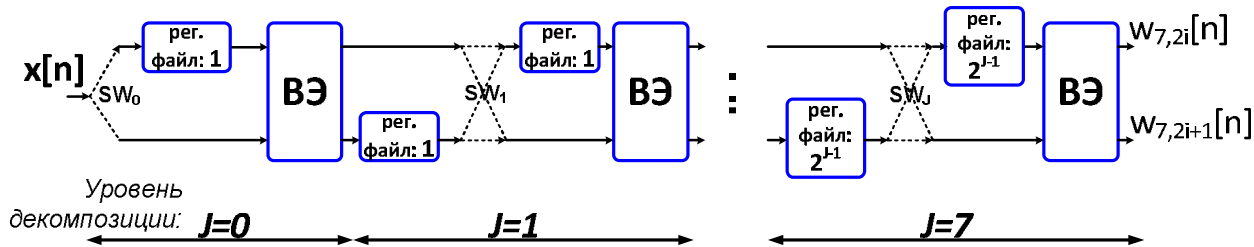


Рис. 17.18 – Поточная архитектура процессора ПДВП с числом уровней декомпозиции, равным 8.

Данная архитектура позволяет вычислять полное восьмиуровневое ПДВП с использованием одного вычислительного элемента на каждом уровне. Для правильного обмена данными между ВЭ на смежных уровнях вводятся регистровые файлы, работающие по принципу FIFO, а также коммутаторы SW_J . Работа данных блоков сводится к формированию корректной последовательности данных, поступающих на вход ВЭ на каждом уровне в соответствии с сигнальным ориентированным графом.

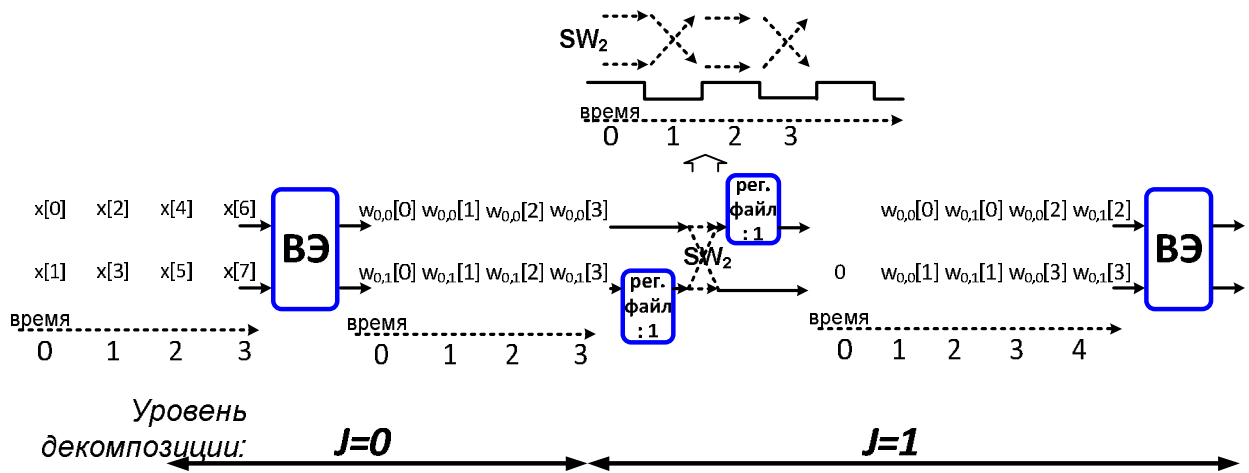


Рис. 17.19 – Граф потока данных между первым и вторым уровнем процессора ПДВП.

Ниже приведен пример графа потока данных на входе и выходе ВЭ первого уровня, и далее сформированный поток данных, поступающих на вход ВЭ второго уровня декомпозиции (см. рис. 17.19). Коммутатор SW_2 в четные отсчеты времени должен замыкать линии связи в прямом направлении, а в нечетные отсчеты времени в перекрестном.

На каждом последующем уровне декомпозиции коммутатор SW_J ($J = 2..7$) переключается с периодичностью в два раза большей, чем коммутатор SW_{J-1} на предыдущем уровне. Что в итоге позволяет с учетом введения регистровых файлов обеспечить правильное формирование потока данных.

Таким образом, рассмотренная поточная структура может быть легко масштабируема для любого числа уровней декомпозиции дерева ПДВП, при этом она не требует ввода сложной схемы управления.

17.8 Операционный блок процессора

Под операционным блоком в данной работе понимается устройство, функционально реализующее двухканальный банк фильтров анализа на основе лестничных структур. При этом данный блок должен удовлетворять следующим требованиям:

- полное функциональное соответствие математической модели на арифметике с фиксированной запятой;
- минимизация аппаратных затрат;
- обеспечение высокой пропускной способности;
- параметризация составных элементов блока с целью ускорения процесса проектирования при изменении требований разработчика.

Последний критерий подразумевает под собой то, что результатом разработки должна стать параметризованная модель процессора ПДВП, в которой разработчик может изменять следующие характеристики: конкретный тип лестничной структуры (число шагов лестничной структуры, значения коэффициентов b_0 , b_1 и др.); возможность изменения разрядности основных операционных узлов и регистров для хранения входных, выходных и промежуточных данных, участвующих в вычислениях.

В результате анализа алгоритма предлагается структура операционного блока, изображенная на рис. 17.20.

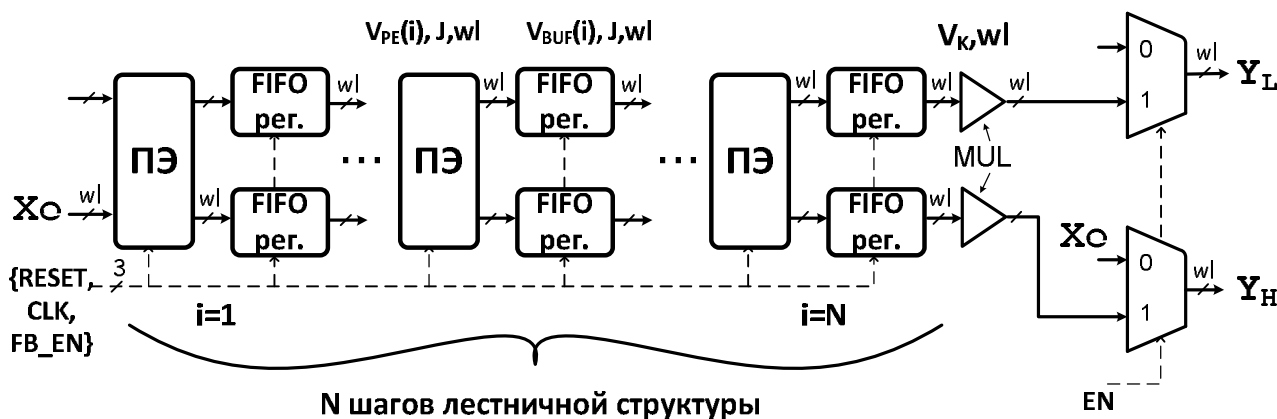


Рис. 17.20 – Структурная схема операционного блока.

Вся структура данного блока задается следующими параметрами:

- wl – разрядность данных, поступающих на вход и соответственно разрядность всех арифметических операций и буферов для хранения промежуточных результатов;
- J – индекс, задающий уровень декомпозиции, на котором используется данный операционный блок (банк фильтров);
- N – число элементарных шагов лестничной структуры;

- $V_{PE}(i) \rightarrow \{id, type, b_0, ind_x, sh_{x1}, sh_{x2}, [sh_{s2}, b_1, delayed]\}, i = 1..N$ – массив из N векторов, каждый из которых задает параметры i -го элементарного шага лестничной структуры (сами параметры будут рассмотрены ниже);

- $V_{BUF}(i) \rightarrow \{d_{xe}, d_{xo}\}, i = 1..N$ – массив из N векторов, каждый из которых задает число элементов задержки d_{xe}, d_{xo} соответственно в верхнем и нижнем каналах после i -го элементарного шага;

- $V_K \rightarrow \{K_1, K_2, n_{ye}, n_{yo}\}$ – вектор, задающий значения квантованных коэффициентов K_1 и K_2 для операции умножения на матрицу \mathbf{K} (17.43), а также номера старших бит n_{ye}, n_{yo} , начиная с которых будут выданы результаты перемножения сигналов на коэффициенты K_1 и K_2 соответственно.

Рассмотрим подробнее структуру операционного блока. В табл. 17.3 приведены входные выходные сигналы данного блока.

Табл. 17.3 – Внешние сигналы операционного блока.

Название	Тип	Разрядность	Описание
<i>FB_EN</i>	вход	1 бит	Сигнал разрешения работы арифметических блоков и блоков хранения промежуточных данных. Если равен 1, значит блоки активны, если равен 0, то в данный момент времени операционный блок не используется, при этом на выходные значения Y_L, Y_H передаются входные значения X_e, X_o .
<i>RESET</i>	вход	1 бит	Сигнал сброса в нулевое значение всех элементов памяти.
<i>CLK</i>	вход	1 бит	Сигнал тактирования элементов памяти
X_e	вход	w_l бит	Входная шина данных
X_o	вход	w_l бит	Входная шина данных
Y_L	выход	w_l бит	Выходная шина данных
Y_H	выход	w_l бит	Выходная шина данных

Принцип работы операционного блока заключается в следующем. На входные шины X_e и X_o поступают четный и нечетный отсчеты входной последовательности, после чего начинается процесс вычисления вейвлет коэффициентов. Результат высчитывается в процессорных элементах ПЭ на основе текущих входных значений и предыдущих значений, хранящихся в буферных блоках. Данные блоки представляют собой регистровые файлы, работающие по принципу FIFO. Число регистров в каждом таком блоке определяется как $n_{xe} \cdot 2^J$ (либо $n_{xo} \cdot 2^J$), где значения n_{xe}, n_{xo} определяют задержку сигнала, соответствующую задержке $z^{-n_{xe}}, z^{-n_{xo}}$ в исходном алгоритме, а значение 2^J определяет число тактов между приходом входных отсчетов из одной субполосы. Данный прием, предложенный в [14,15], необходим для того чтобы значения предыстории на выходах блоков FIFO относились к той же субполосе, что и текущие входные сигналы.

Пара результатов, вычисленная в последнем элементе ПЭ, умножается на константы K_1 и K_2 , и далее подаётся на выход операционного блока.

Как уже говорилось раньше, для параметризации i -го процессорного элемента ПЭ формируется специальный вектор $V_{PE}(i)$. Ниже даны разъяснения для всех параметров, включенных в данный вектор.

Параметр id (identifier). Идентификатор, задающий один из двух типов процессорного элемента для реализации элементарного шага лестничной структуры: $id = 'PE0'$ при значении коэффициента $b_1^i = 0$ (рис. 17.21), в противном случае $id = 'PE1'$ (рис. 17.22).

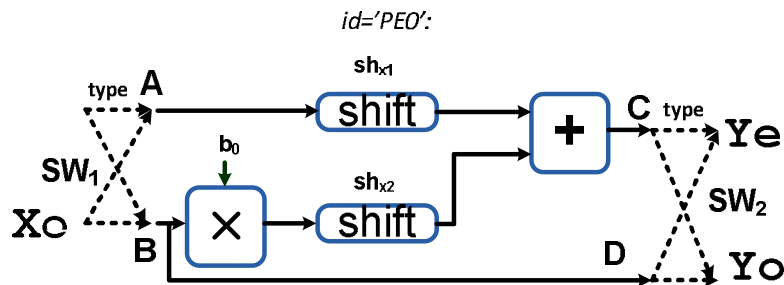


Рис. 17.21 – Структурная схема процессорного элемента при $id = 'PE0'$.

Параметр $type$ задает тип элементарного шага, реализуемого ПЭ: $type = 's'$ если это прямой шаг (17.11), и $type = 't'$ если шаг дуальный (17.12).

Параметры b_0, b_1 могут принимать значения от «минус» 2^{wl-1} до «плюс» $(2^{wl-1} - 1)$ и соответствуют значениям mb_0, mb_1 для квантованных коэффициентов b_0^i, b_1^i . В случае $id = 'PE0'$ параметр b_1 игнорируется.

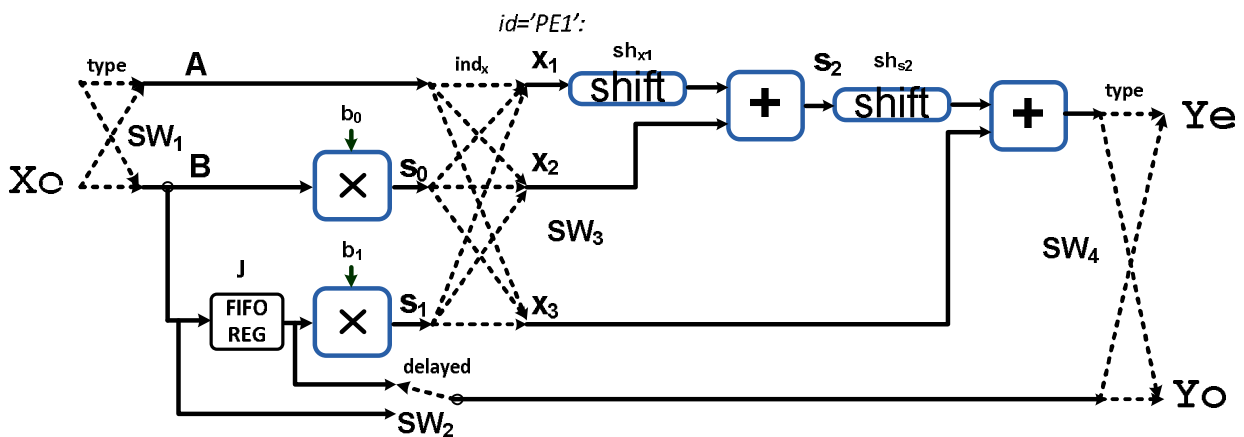


Рис. 17.22 – Структурная схема процессорного элемента при $id = 'PE1'$.

Параметр ind_x представляет собой массив из трех чисел, принимающих значения 1,2, либо 3. Данный параметр определяет очередность суммирования сигналов A, s_1, s_2 для ПЭ с $id = 'PE1'$. К примеру, если параметр $ind_x = [2\ 3\ 1]$, то для получения выходного результата сначала выполняется операция $A + s_2$ и далее к получившемуся результату прибавляется значение s_1 .

Параметры $sh_{x1}, sh_{x2}, sh_{s2}$ определяют число арифметических сдвигов для внутренних сигналов процессорного элемента x_1, x_2 и s_2 соответственно.

Параметр *delayed*. Для того чтобы разъяснить назначение данного параметра, рассмотрим участок исходного алгоритма (рис. 17.23), на котором сначала выполняется прямой шаг $s(z) = b_0 + b_1z^{-1}$ и далее следует задержка сигнала z^{-1} . Как видно из рисунка, сигнал $Y_e(n - 1)$, который должен быть на выходе всей схемы, уже сформирован внутри процессорного элемента. Поэтому при значении *delayed* = 1 на выход ПЭ выдается не сигнал $Y_e(n)$, а сигнал, $Y_e(n - 1)$, что позволяет сэкономить для данной схемы на J -ом уровне декомпозиции 2^J регистров.

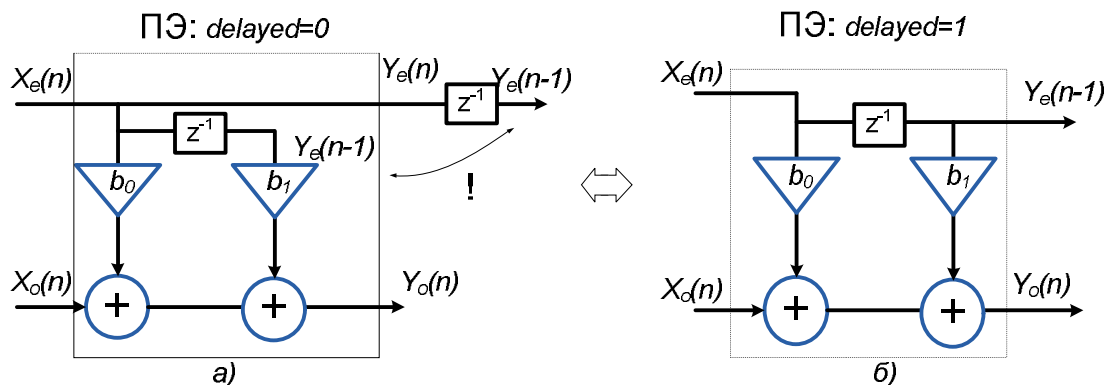


Рис. 17.23 – Структура схема процессорного элемента PE при *delayed* = 0 (а) и *delayed* = 1 (б).

Отметим, что блоки арифметического сдвига *shift*, а также коммутаторы SW_1, SW_2, SW_3, SW_4 не используют аппаратного ресурса системы. Операция арифметического сдвига вправо сводится к смещению сигнальных линий перед подачей их на вход сумматора. А указанные выше коммутаторы в соответствии с параметрами *type, ind_x, delayed* реализуются в виде замыкания соответствующих линий.

Таким образом, процессорный элемент для случая $id = 'PE0'$ в конечном итоге будет реализован на одном сумматоре и одном умножителе, а процессорный элемент для $id = 'PE1'$ – на двух сумматорах, двух умножителях, и регистровом файле, состоящем из 2^J регистров. Все используемые элементы имеют разрядность равную wl .

17.9 Список использованных источников

1. Воробьев В.И., Грибунин В.Г., Теория и практика вейвлет преобразования, С-П., ВУС, 1999, 204 с.
2. Bellanger M., Bonnerot G., Coudreus M., Digital filtering polyphase network: application to sample rate alteration and filter banks, IEEE Trans. Acoust., Speech and signal processing, V ASSP-24 Apr. 1976, pp. 109-114.
3. Sweldens W, The lifting scheme: A custom-design construction of biorthogonal wavelets, Appl. Comput. Harmon. Anal, 1996, pp. 186-200.
4. Daubechies I., Sweldens W., Factoring wavelet transforms into lifting steps, Journal of Fourier Anal. Appl., Vol. 4, No. 3, 1998, pp. 247-269.

-
5. Bultheel A., Van Barel M., Linear algebra, rational approximation and orthogonal polynomials, North-Holland, 1997, 446 p.
 6. Родионов М.М., Петровский А.А., FPGA-реализация двухканального банка фильтров на лестничных структурах, Материалы 5-ой международной конференции-форума «Информационные системы и технологии IST'2009» в 2 Ч., Мн. 2009, с. 232-235.
 7. Coors M., and etc., Design and DSP implementation of fixed-point systems, EURASIP journal on applied signal processing, 2002, No.9, pp. 908-925.
 8. Menard D., Chillet D., Sentieys O., Floating-to-fixed-point conversion for digital signal processors EURASIP journal on applied signal processing, 2006, pp. 1-19.
 9. Teans M.A., Lopez J., Zapata E.L., A configurable architecture for wavelet packet transform, VLSI signal process, Vol. 32, 2002, pp. 255-273.
 10. Arguello F., and etc. F., Architecture for wavelet packet transform based on lifting steps, Journal parallel computation, Vol. 28, No. 7-8, 2002, pp. 1023-1037.
 11. Petrovsky Al., Petrovsky A., Audio coding with a masking threshold adapted wavelet packet based on run-time reconfigurable processor architecture, The proc. of 110th AES convention, convention paper 5298, May 12-15 2001, Amsterdam, Netherlands, 8 pages.
 12. Petrovsky Al., Petrovsky A., Reconfigurable processing: the solution to audio coding in real-time, The proc. of the 9th International conference "Mixed design of integrated circuits and systems", MIXDES'2002, 20-22 June 2002, Wroclaw, Poland, pp.629-634.
 13. Petrovsky Al., Petrovsky A., Dynamic algorithm transforms for reconfigurable real-time audio coding processor, The proc. of the Intern. Conference on Parallel Computing in Electrical Engineering, PARELEC'02, Warsaw, Poland, Sep. 22-25, 2002, IEEE Computer Society Press, Los Alamitos, California, 2002, pp. 422-424.
 14. Petrovsky Al., Krahe D., Petrovsky A., Real-time wavelet packet-based low bit rate audio coding on a dynamic reconfiguration system, AES, Convention paper 5778, 114th Convention, 2003 March 22-25, Amsterdam, The Netherlands, 22 pages.
 15. Wang Ch., Gang W.S., Efficient VLSI architecture for lifting-based discrete wavelet packet transform, IEEE transactions on circuit and systems, Vol. 54, No. 5, 2007, pp. 422-426.

18 ВЫЧИСЛИТЕЛЬНАЯ ПЛАТФОРМА НА ОСНОВЕ ЦПОС TMS320C6713 И РЕКОНФИГУРИРУЕМОЙ ЛОГИКИ ДЛЯ МУЛЬТИМЕДИА ПРИЛОЖЕНИЙ РЕАЛЬНОГО ВРЕМЕНИ

Лившиц М.З., Петровский Ал.А., Качинский М.В., Станкевич А.В., Петровский А.А.

АННОТАЦИЯ

В главе обсуждаются особенности аппаратной реализации двухпроцессорной вычислительной платформы на основе ЦПОС TMS320C6713 и периферийного процессора, реализуемого на реконфигурируемой логике, для мультимедиа приложений реального времени, представлены характеристики аппаратной платформы. Показана реализация двух систем компрессии речи и звука: комбинированной системы кодирования широкополосных речевых сигналов и перцептуальный кодер звука на базе ПДВП процессора с динамической трансформацией частотно-временного плана. Приведены результаты субъективных и объективных оценок качества восстановленных сигналов.

Ключевые слова – FPGA, VLIW, WDFТ, ЦПОС, ПДВП процессор.

СОДЕРЖАНИЕ

18.1	Вычислительная платформа для мультимедиа приложений реального времени.....	398
18.1.1	Краткое описание ЦПОС TMS320C6713.	398
18.1.2	Структура вычислительной платформы.....	399
18.1.3	Организации памяти, инициализация аппаратуры	401
18.1.4	Функциональный модуль для задач цифровой передачи звуковых сигналов	402
18.2	Система кодирования широкополосных речевых сигналов с препроцессором редактирования шума.....	403
18.2.1	Структура комбинированной системы кодирования широкополосной речи со встроенным препроцессором подавления шума.....	403
18.2.2	Организация вычислительного процесса.....	410
18.2.3	Оценки качества системы компрессии	413
18.2.4	Объективная и субъективная оценка качества препроцессора редактирования шума.....	415
18.2.5	Эффективность комбинированной системы кодирования звуковых сигналов	418
18.3	Реализация кодера звука на базе ПДВП с динамической трансформацией частотно-временного плана на вычислительной платформе в реальном времени.....	420
18.3.1	Структура ПДВП-кодера звука.....	420
18.3.2	Организация вычислительного процесса.....	421
18.3.3	Оценка эффективности кодера.....	423
18.4	Список использованных источников.....	426

18.1 Вычислительная платформа для мультимедиа приложений реального времени

18.1.1 Краткое описание ЦПОС TMS320C6713

ЦПОС TMS320C6713 [1] представляет собой 32-битный процессор с плавающей точкой и обеспечивающий около 1350 MFLOPS или 1800 MIPS при тактовой частоте 225 МГц. Процессор основан на расширенной («продвинутой») VLIW-архитектуре, разработанной Texas Instruments Inc. (TI) Его восемь независимых функциональных устройств и 256-битная VLIW-архитектура позволяет за один такт обрабатывать до восьми 32-битных инструкций. Процессор имеет широкий набор периферийных устройств, оптимизированный для обработки аудиосигналов; процессорное ядро включает 8 независимых функциональных блоков: 2 целочисленных АЛУ, 4 АЛУ с плавающей запятой, 2 умножителя; архитектура «загрузка / сохранение» с 32-мя 32-разрядными регистрами общего назначения (РОН); двухуровневая архитектура памяти: 4 Кбайт кэш программ первого уровня L1P (прямого отображения), 4 Кбайт кэш данных первого уровня L1D (наборно-ассоциативный на 2 вхождения), 256 Кбайт память второго уровня L2 (64 Кбайт кэш/ОЗУ и 192 Кбайт ОЗУ); 16/32-разрядный интерфейс внешней памяти (EMIF), позволяющий адресовать 512 Мбайт адресного пространства.

Периферия включает два последовательных многоканальных аудиопорта (McASPs), два последовательных многоканальных буферизированных порта (McBSPs), две шины I2C, один модуль ввода/вывода общего назначения (GPIO), два таймера общего назначения, интерфейс хост порта (HPI) и интерфейс для подключения внешней памяти (SDRAM, SBRAM и периферийного асинхронного оборудования).

В качестве интегрированной среды разработки использовалось ПО фирмы Texas Instruments Code Composer Studio v3.1® (рис. 18.1).

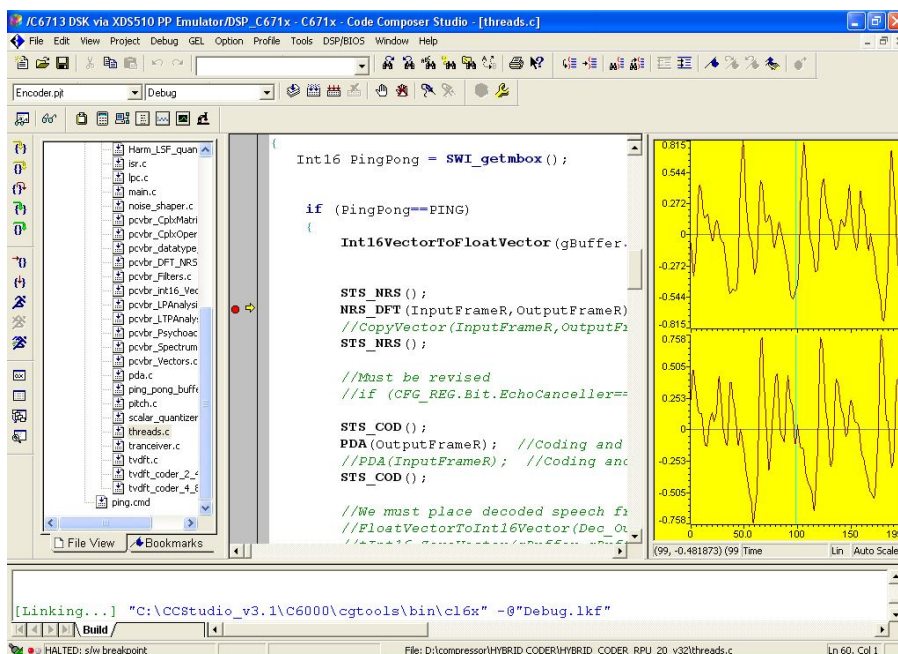


Рис. 18.1 – Интегрированная среда разработки и отладки Code Composer Studio v3.1®

18.1.2 Структура вычислительной платформы

Платформа [2], построена на базе двухпроцессорного ядра, в состав которого входят ЦПОС TMS320C6713B [1] и ПЛИС Xilinx семейства Spartan-3 XC3S200 со структурой FPGA [3], выполняющая функции периферийного процессора (структурная схема представлена на рис. 18.2).

ЦПОС обеспечивает выполнение сложных алгоритмов цифровой обработки аудиосигналов. Периферийный процессор предназначен для реализации внешнего интерфейса аппаратной платформы, предварительной и окончательной обработки данных, обеспечения требуемых форматов данных интерфейсов, формирования и разборки информационных пакетов внешнего интерфейса, а также для реализации алгоритмов цифровой обработки сигналов с целью разгрузки ЦПОС.

Для реализации внешнего интерфейса системы используются порты ввода/вывода FPGA (для использованного типа корпуса PQ208 число пользовательских выводов – 141). ПЛИС семейства Spartan-3 поддерживают 26 стандартов ввода/вывода, из которых 8 являются дифференциальными (LVDS, LVPECL и др.). Для возможности подключения к системам с уровнями напряжений более 3.3 В для части портов ввода/вывода предусмотрены преобразователи уровня. Дополнительные возможности по сопряжению предоставляют триггеры Шмитта, позволяющие устойчиво работать с сигналами, имеющими длинные фронты нарастания и спада. Порты ввода/вывода могут быть использованы как одна или несколько параллельных шин данных нужной разрядности, а также как линии для последовательной передачи данных.

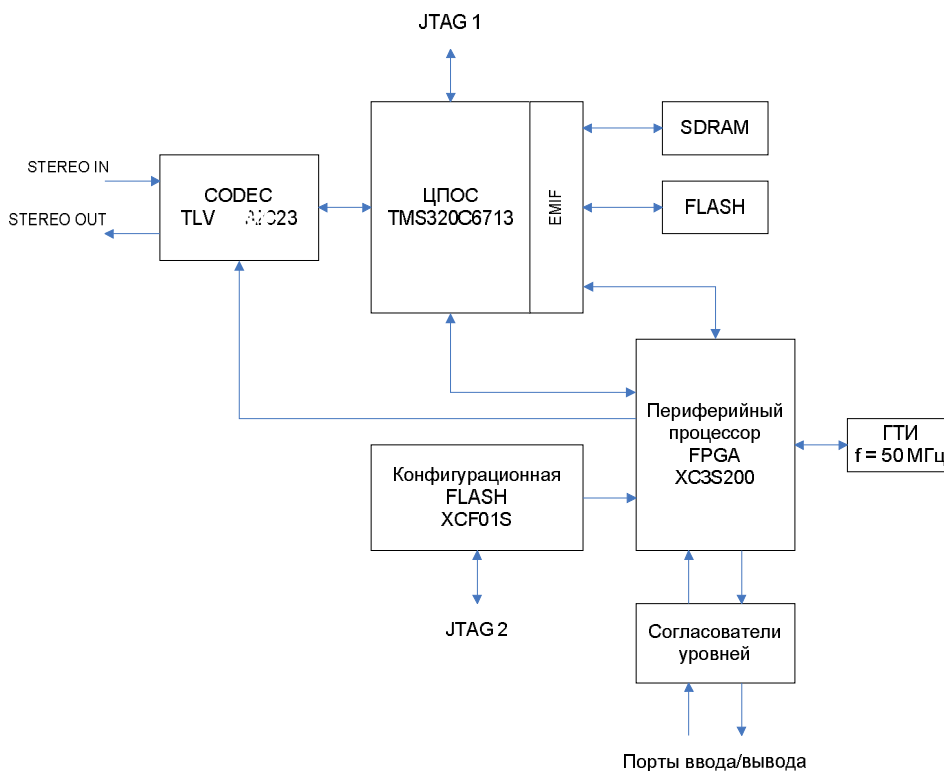


Рис. 18.2 – Структурная схема платформы для мультимедиа приложений реального времени

FPGA помимо логических ячеек содержит триггеры, блоки памяти, матричные умножители и схемы цифровой подстройки частоты. При реализации функций цифровой обработки

сигналов на периферийном процессоре могут использоваться 12 блоков матричных умножителей FPGA, имеющих разрядность операндов 18 бит и результата 36 бит.

Разработка проекта периферийного процессора может быть реализована с помощью САПР Xilinx ISE® [4]. В этом случае конфигурационная последовательность FPGA генерируется с помощью утилиты iMPACT пакета и загружается через интерфейс JTAG в конфигурационную флэш-память платформы. Каждый раз после включения питания FPGA считывает эту последовательность из флэш-памяти.

Аналоговый тракт разработанной платформы выполнен на базе высокоэффективного стерео аудиокодека TLV320AIC23B [5] имеющего следующие характеристики:

- 1) многобитный сигма-дельта АЦП ($SNR = 90$ дБ);
- 2) многобитный сигма-дельта ЦАП ($SNR = 100$ дБ);
- 3) цифровой источник питания 1.42 В – 3.6 В;
- 4) буферный и аналоговый источник 2.7 В – 3.6 В;
- 5) частота дискретизации 8 кГц – 96 кГц;
- 6) программный контроль через McBSP-совместимый мультипротокольный последовательный порт;
- 7) ввод/вывод аудиоданных через McBSP-совместимый программируемый интерфейс;
- 8) I2C-совместимый интерфейс, требующий только один McBSP как для АЦП, так и для ЦАП;
- 9) входы стерео линии (встроенный программируемый усилитель, аналоговый обводной путь кодека, мультиплексируемые линии линейного стерео входа и микрофонной линии);
- 10) линейный стерео выход;
- 11) гибкое управление потребляемой мощностью осуществляется программно (потребляемая мощность в режиме проигрывания 23 мВт; в режиме готовности <150 мкВт; в режиме пониженного энергопотребления <15 мкВт).



Рис. 18.3 – Вычислительная платформа с реконфигурируемой логикой для мультимедиа приложений реального времени

Данное архитектурное решение позволяет получить следующие особенности аппаратной платформы:

- 1) мощное вычислительное ядро с плавающей запятой для обработки данных с производительностью до 1350/1800 MFLOPS/MIPS;
- 2) возможность распараллеливания задач обработки в двухпроцессорной системе;
- 3) реконфигурируемость платформы путем изменения конфигурационной последовательности FPGA и программ ЦПОС;
- 4) возможность встраивания платформы в разнообразное оборудование за счет гибкой настройки внешнего интерфейса с помощью периферийного процессора;
- 5) использование стандартных средств конфигурирования и программирования.

Основные электрические параметры платформы:

- 1) напряжение питания модуля +5 В постоянного тока;
- 2) ток потребления – не более 300 мА;
- 3) потребляемая мощность зависит от загрузки ЦПОС и ПЛИС (от реализуемых алгоритмов) и составляет не более 1.5 Вт.

Фотография вычислительной платформы представлена на рис. 18.3.

18.1.3 Организации памяти, инициализация аппаратуры

Загрузчик. Настройка генератора тактовых импульсов осуществляется сразу после включения питания ЦПОС и вывода его из состояния сброса. При первоначальном сбросе ЦПОС, происходит автоматическое копирование содержимого флэш-памяти (от англ. – «FLASH») из адресного пространства CE1 объемом 1 Кбайт. В данном диапазоне располагается так называемый загрузчик (от англ. – «boot loader»). Загрузчик реализует функции инициализации аппаратуры ЦПОС и его периферийных устройств, а также начальную инициализацию внутренней и внешней памяти (копирование кода исполняемой программы и данных, необходимых для ее функционирования) [6]. Карта памяти вычислительной платформы представлена на рис. 18.4.

Настройка тактовых частот ЦПОС и периферии. По схемотехническому решению ЦПОС тактируется внешним тактовым генератором с частотой 50 МГц. Перед настройкой оборудования в загрузчике осуществляется конфигурирование внутреннего программируемого генератора с автоподстройкой частоты (PLL-генератор) (от англ. – «Phase-Locked Loop Generator» – PLL-generator). Инициализация производится в строго определенной последовательности, которая имеет свои особенности для каждого семейства ЦПОС [7]. С помощью PLL-генератора осуществляется тактирование, как самого ЦПОС так и его периферийного оборудования. Существует возможность формирования тактовых импульсов с различными частотами следования отдельно для ЦПОС и периферийных устройств.

В данной реализации PLL-генератор настроен следующим образом: тактовая частота ЦПОС составляет 200 МГц, частота интерфейса внешней памяти (от англ. – «External Memory Interface» – EMIF) составляет 100 МГц.

Интерфейс внешней памяти. Настройка интерфейса внешней памяти осуществляется в загрузчике путем конфигурирования регистров всех адресных пространств CE0–CE3, в регистр

таймера обновления SDRAM заносится константа пересчета соответствующая 15.6 мкс (интервал обновления данных для динамической памяти) [8].

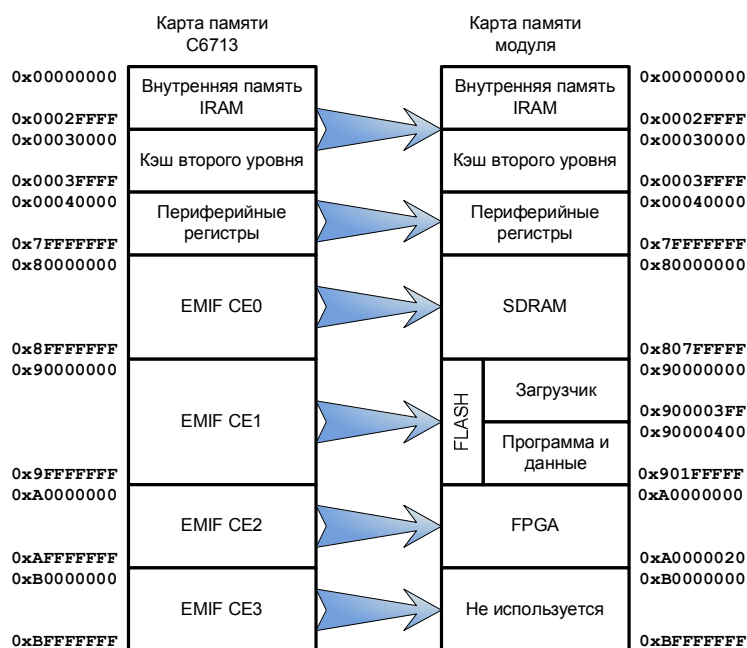


Рис. 18.4 – Карта памяти вычислительной платформы

Копирование исполняемой программы и данных. После настройки PLL-генератора и EMIF осуществляется копирование кода программы и данных в соответствии с картой памяти указанной линковщику на этапе компиляции. Затем управление передается на точку входа (процедуру main).

Инициализация и запуск программы. После получения управления процедура инициализации осуществляет следующие действия:

- 1) настройка переменных окружения среды;
- 2) инициализация стека;
- 3) настройка кэш-памяти второго уровня;
- 4) инициализация констант и кодовых книг;
- 5) конфигурирование обработчиков прерываний;
- 6) настройка расширенного контроллера прямого доступа к памяти (от англ. – «Enhanced Direct Memory Access» – EDMA);
- 7) настройка многоканальных буферизированных портов (от англ. – «Multichannel Buffered Serial Port» – McBSP) для обмена с аудиокодеком;
- 8) настройка аудиокодека и его запуск;
- 9) разрешение обработчиков настроенных прерываний;
- 10) инициализация операционной системы DSP/BIOS;
- 11) глобальное разрешение обработки прерываний.

18.1.4 Функциональный модуль для задач цифровой передачи звуковых сигналов

Обобщенная функциональная схема модуля представлена на рис. 18.5. Входной и выходной аналоговый тракты построены на базе микросхемы TI TLV320AIC23B [5]. Входные дан-

ные оцифровываются с частотой дискретизации 16 кГц или 44,1 кГц, с аналогичной частотой осуществляется выдача выходных данных или реконструированного речевого сигнала. Конфигурирование аудиокодека осуществляется через McBSP0, а прием и выдача звуковых отсчетов – через McBSP1. Аппаратная реализация алгоритмов кодера была осуществлена согласно описанию, представленному в [9,10].

Для организации ввода-вывода речевой информации в реальном масштабе времени используется техника двойной буферизации известная также как PING-PONG буфер, который реализуется при помощи контроллера EDMA и так называемой техники связанных передач (от англ. – «Linked Transfers») [11].

Из-за невозможности поместить весь код и данные широкополосного компрессора во внутренней памяти ЦПОС, наиболее критичные ко времени исполнения процедуры и функции алгоритмов кодирования размещены во внутренней памяти ЦПОС. Остальные данные и код программы перенесены во внешнее ОЗУ. Для повышения скорости доступа к внешнему ОЗУ используются встроенные возможности кэширования данного ЦПОС.

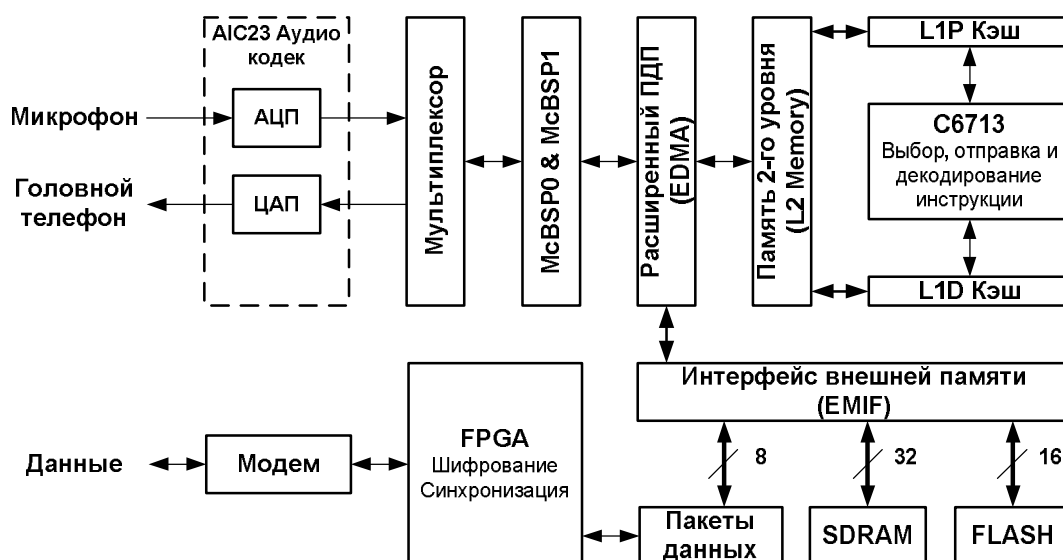


Рис. 18.5 – Функциональная схема модуля для задач цифровой передачи звуковых сигналов

Диаграмма состояний вычислительного процесса с распределением приоритетов между аппаратными и программными прерываниями, а также фоновыми задачами представлена на рис. 18.6.

18.2 Система кодирования широкополосных речевых сигналов с пре-процессором редактирования шума

18.2.1 Структура комбинированной системы кодирования широкополосной речи со встроенным препроцессором подавления шума

Так как в описываемой системе компрессии широкополосного речевого сигнала уже выбран тип частотно-временного преобразования и используется психоакустическая модель для управления реконфигурацией структуры банка кодовых книг, то с целью экономии вычис-

лительного ресурса и достижения высокого качества комбинированной системы необходимо использовать второй подход (обе системы используют общие ресурсы: частотно-временное преобразование, психоакустическая модель и т.п.). Таким образом, общая структура комбинированной системы кодирования широкополосной речи может быть представлена в виде схемы, изображенной на рис. 18.7.

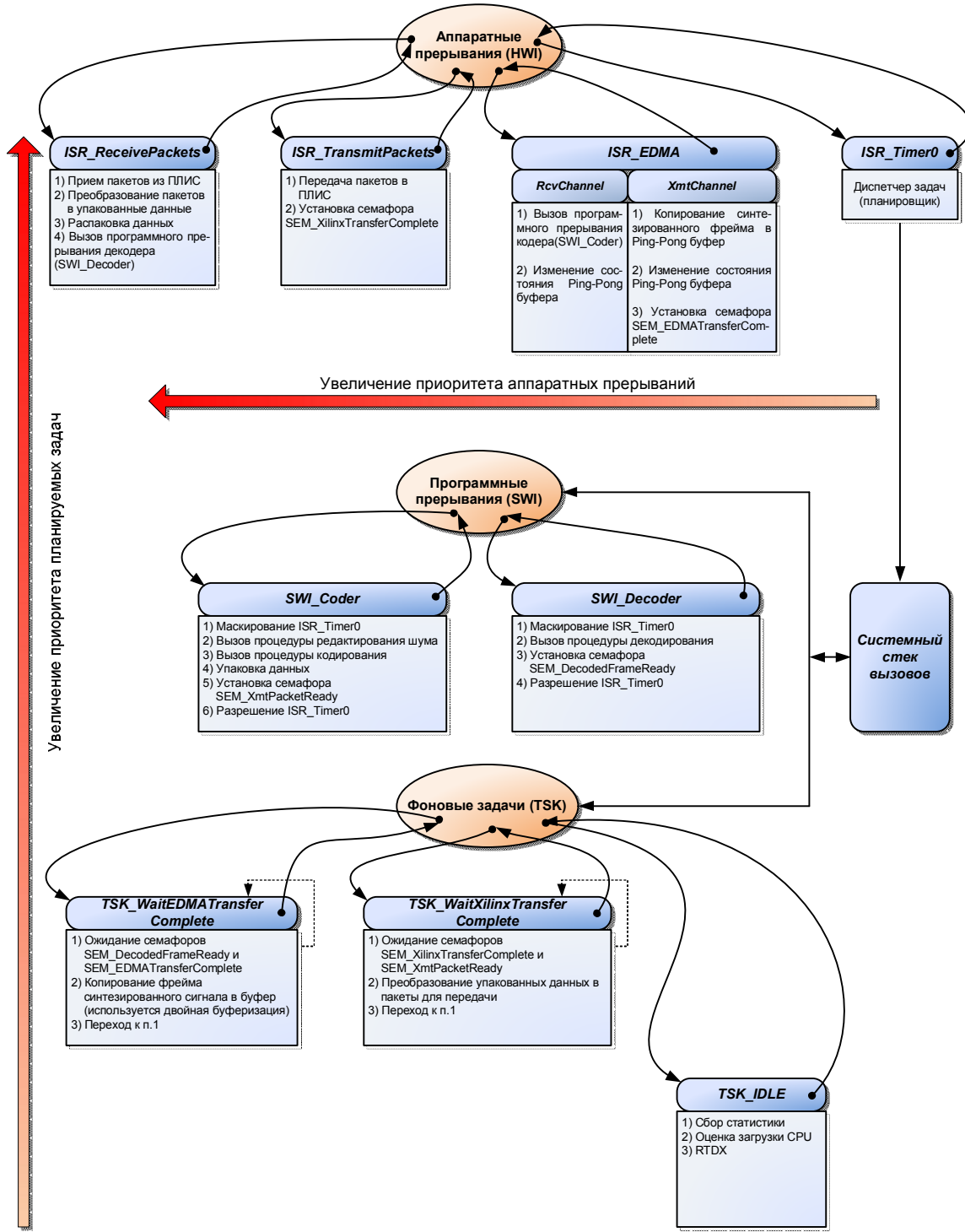


Рис. 18.6 – Диаграмма состояний (диспетчеризация задач)

Применение дискретного преобразования Фурье с неравномерным частотным разрешением (Warped DFT (WDFT)) в перцептуальной системе редактирования шума не только как базиса для определения модели маскирования, но и как инструмента спектрального анализа,

позволит добиться наилучшего качества по сравнению с обычными системами повышения качества речи. Это объясняется тем, что весь процесс обработки осуществляется в единой перцептуальной области (нет необходимости в преобразованиях между различными частотными шкалами). Более того, обработка речевого сигнала, осуществляемая в критических частотных полосах, более точна в контексте психоакустического моделирования [12,13].

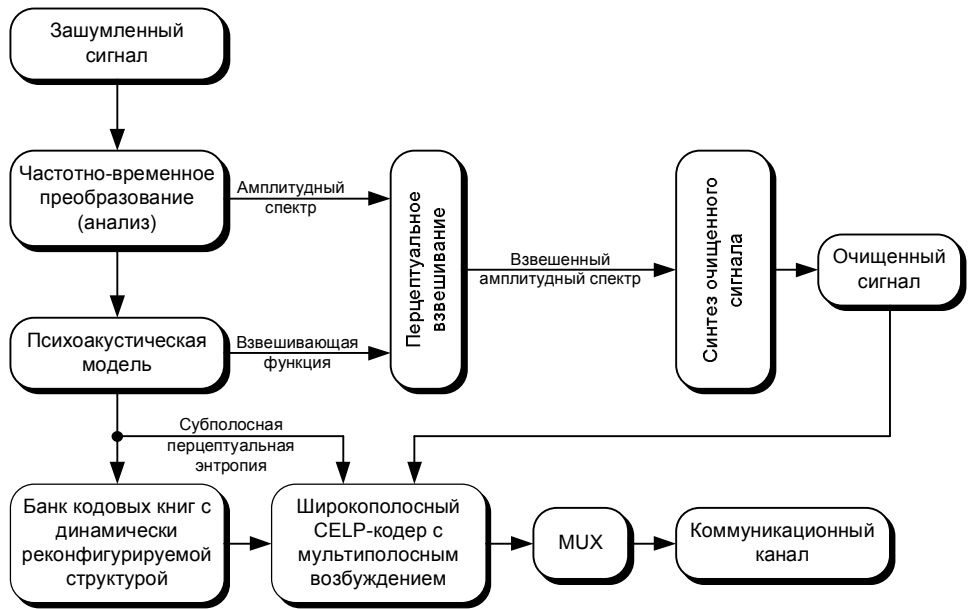
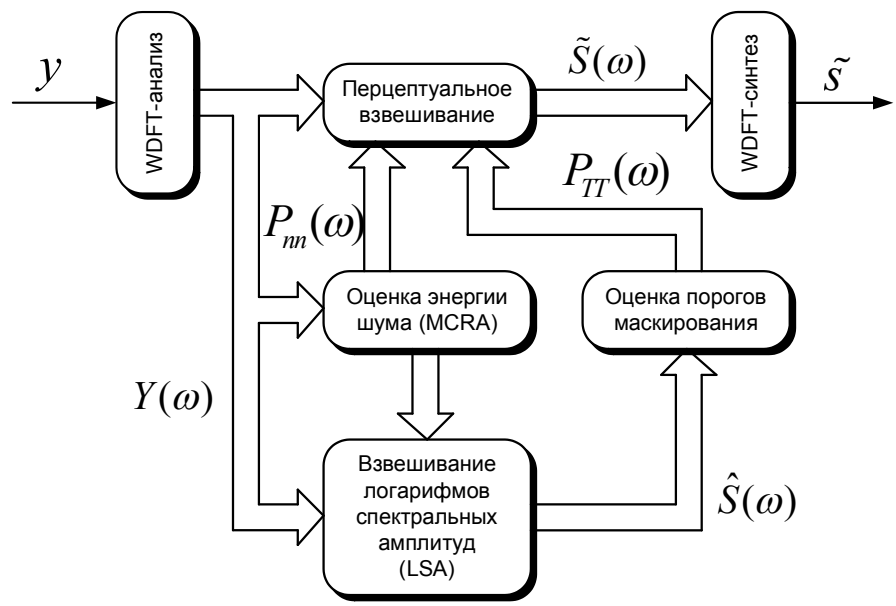


Рис. 18.7 – Структура комбинированной системы редактирования шума и кодирования широкополосного речевого сигнала



y – входной сигнал; $Y(\omega)$ – СПМ входного зашумленного сигнала; $P_{nn}(\omega)$ – СПМ шума; $P_{TT}(\omega)$ – оценка энергии порога маскирования; $\tilde{S}(\omega)$ – СПМ очищенного сигнала; $\hat{S}(\omega)$ – взвешенная оценка СПМ зашумленного сигнала; \tilde{s} – выходной очищенный сигнал.

Рис. 18.8 – Структура математической модели перцептуальной системы подавления шума на базе WDFT [14]

Уточненная структурная схема препроцессора повышения качества речи на базе WDFT представлена на рис. 18.8.

Алгоритм основан на хорошо известном правиле спектрального взвешивания [15]:

$$H^{JND}(\omega) = \min \left\{ 1, \sqrt{\frac{P_{TT}(\omega)}{P_{nn}(\omega)} + \xi_{resnoise}} \right\}, \quad (18.1)$$

где JND означает неслышимое шумовое искажение.

Легко заметить, что, если энергия остаточного шума лежит ниже порога маскирования, выражение под корнем больше единицы и речевой сигнал не искажается так как $H^{JND}(\omega) = 1$. В противном случае, окружающий шум оптимально ослабляется до невоспринимаемого уровня. Для расчета взвешивающих коэффициентов необходима оценка порога маскирования и СПМ шума.

Комбинирующая система кодирования широкополосной речевой информации с встроенным препроцессором повышения качества речи может быть представлена в виде структурной схемы, изображенной на рис. 18.9.

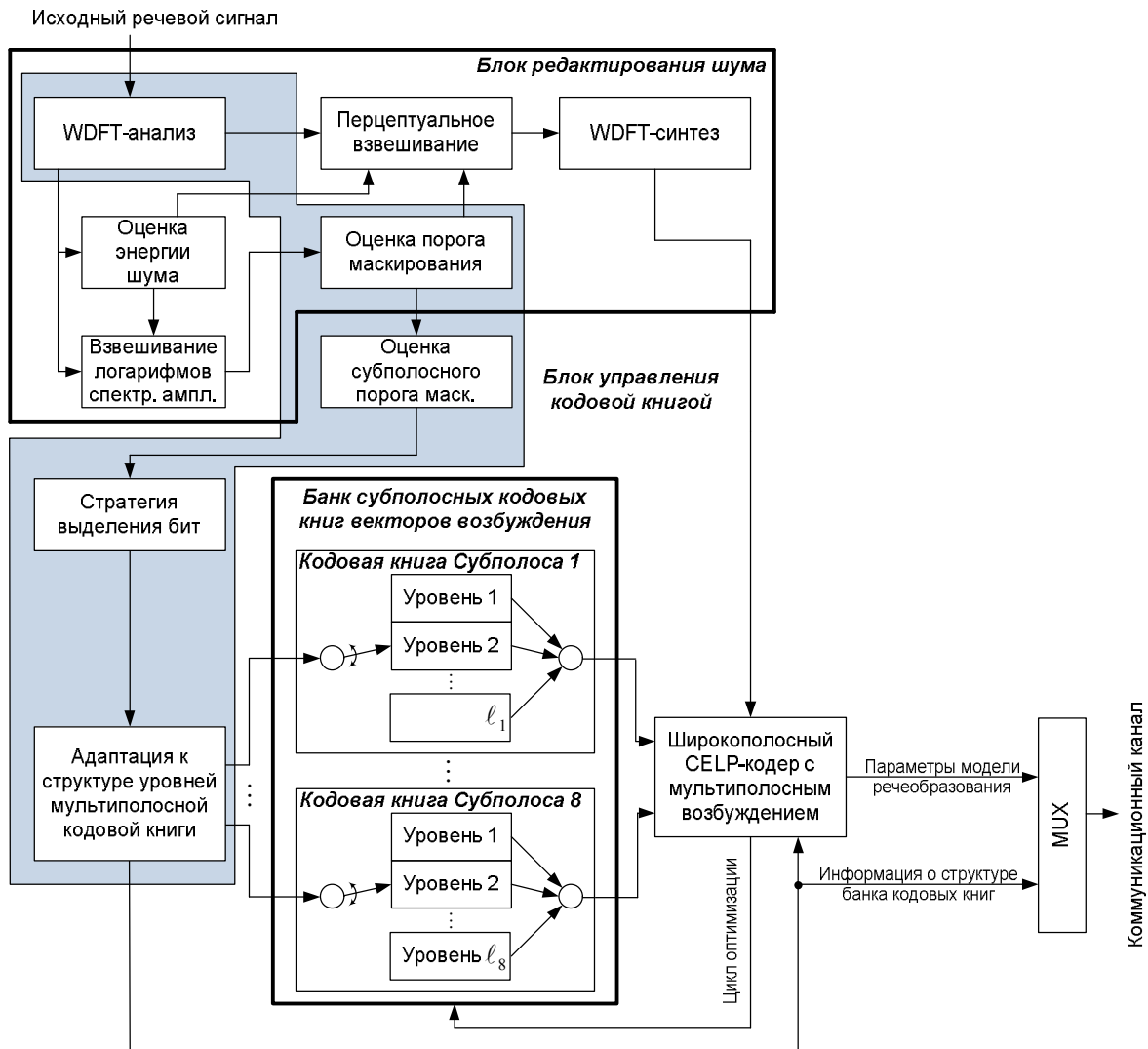


Рис. 18.9 – Комбинирующая система кодирования широкополосного речевого сигнала [16]

Как видно из рис. 18.9 система состоит из четырех основных блоков: блока редактирования шума, блока управления кодовой книгой, банка субполосных кодовых книг векторов возбуждения и непосредственно широкополосного CELP-кодера с мультиполосным возбуждением. Кодер обрабатывает входной сигнал фреймами с длительностью 320 отсчетов, дискретизированных с частотой дискретизации 16 кГц и вносит алгоритмическую задержку 20 мс. Пре-процессор редактирования шума использует обработку фрейма входного сигнала с длительностью равной длительности фрейма широкополосного кодера. При этом анализ зашумленного речевого сигнала осуществляется фреймами с 50% перекрытием. В блоке WDFT-синтеза осуществляется формирование обработанного фрейма сигнала путем перекрытия с накоплением. Более детально алгоритм описан в [14] Таким образом, система редактирования шума обрабатывает два субфрейма на один фрейм кодера. Для наглядности, формирование входных фреймов системы редактирования шума и фреймов кодера представлено на рис. 18.10.

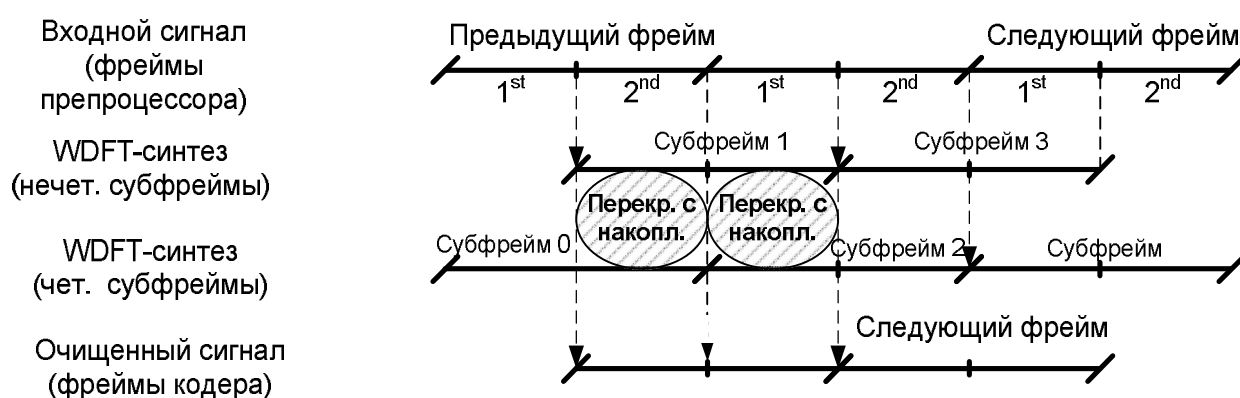


Рис. 18.10 – Формирование фреймов для WDFT-анализа-синтеза

Как видно из рис. 18.10, блок редактирования шума добавляет алгоритмическую задержку, равную половине фрейма кодера (10 мс), что приводит к общей алгоритмической задержке комбинированной системы кодирования речевой информации, равной 30 мс.

Процесс обработки входного сигнала, кодирования и сжатия информации можно представить следующей последовательностью операций:

- дискретизация входного сигнала с частотой 16 кГц, 16 бит/отсчет;
- формирования фрейма входного сигнала длительностью 20 мс (320 отсчетов);
- обработка фрейма исходного сигнала в блоке редактирования шума, получение фрейма очищенного речевого сигнала и значений величины субполосных порогов маскирования [14,17];
- определение структуры банка субполосных кодовых книг векторов возбуждения в блоке управления кодовой книгой на основании оценки субполосных порогов маскирования [16,18] для кодирования текущего фрейма сигнала;
- реконфигурирование банка субполосных кодовых книг векторов возбуждения в соответствии с частотно-временным планом кодируемого фрейма сигнала [17];
- кодирование и квантование параметров модели речеобразования в блоке широкополосного CELP-кодера с мультиполосным возбуждением [17].

Схема квантования параметров модели речеобразования представлена в табл. 18.1.

Табл. 18.1 – Характеристика квантования параметров модели речеобразования

Параметр модели	Кол-во парамет./фрейм	Кол-во бит/параметр	Кол-во бит/фрейм	Кол-во фреймов/с	Скорость потока, бит/с
<i>LSF</i>	16	2	32	50	1600
G_{model}	1	7	7		350
$Delays_{LTP}$	4	7	28		VAR (0–400)
$Gains_{LTP}$	4	5	20		VAR (0–1000)
<i>PPF</i>	1	4	4		200
<i>depth</i>	1 вектор из 8 элементов	1.25	9		450
<i>Excitation Gains(EG)</i>	$EG = 4 \cdot \sum_{b \in Bands} 5$	5	VAR (0 – 160)		VAR (0–8000)
<i>Book Indexes (BI)</i>	$BI = 4 \cdot \sum_{b \in Bands} \log_2(depth(b))$	VAR (4 – 8)	VAR (0 – 224)		VAR (0–11200)
<i>Результирующий поток данных:</i>					VAR (max. 24200)

При расчёте вычислительной сложности не учитывались операции по инициализации регистров, обработке входа и выхода из аппаратных и программных прерываний, и т.п. Из оценки также были исключены действия, необходимые для дискретизации и предварительной фильтрации входного сигнала, которые выполняются на аппаратном уровне с помощью специализированных микросхем аналогового интерфейса без привлечения ЦПОС. Вычислительная сложность тригонометрических операций и трансцендентных функций определялась исходя из предположения о том, что для этого будут использоваться численные методы, реализованные в специализированных библиотеках для ЦПОС [19].

Оценка вычислительного ресурса, необходимого для реализации отдельных частей комбинированной системы, с приблизительной оценкой сложности операций сведена в табл. 18.2 и табл. 18.3.

Расчет вычислительной сложности препроцессора повышения качества речевого сигнала табл. 18.2 и широкополосного компрессора/декомпрессора табл. 18.3 показывает, что при значениях $N = 320$, $M = 512$, $b = 8$, $D_{min} = 32$, $D_{max} = 160$ для компрессии/декомпрессии одного фрейма речевого сигнала с включенной системой редактирования шума требуется около 20 миллионов операций с плавающей запятой (MFLOPS). Так как обработка сигнала осуществляется с периодичностью 50 раз в секунду, то общая вычислительная сложность составит приблизительно 1000 MFLOPS.

В таблицах табл. 18.2, табл. 18.3 приняты следующие обозначения:

N – длина анализируемого фрейма в отсчетах;

M – размер избыточного базиса WDFT [20];

D_{min} – минимальное значение задержки LTP-фильтра ($D_{min} = 32$ отсчета);

D_{max} – максимальное значение задержки LTP-фильтра ($D_{max} = 160$ отсчетов);

B – максимальное количество уровней в субполосной кодовой книге;

L – длина субфрейма ($L = 80$ отсчетов).

Табл. 18.2 – Вычислительная сложность препроцессора повышения качества речи¹

Этап вычислений	Вычислительная сложность, операций	Использование нескольких исполнителей, оптимальная загрузка конвейера	Загрузка двух операндов за такт
Подготовка входных данных	$4N_{mov+mpy}$	частично	да
WDFT-анализ	$(2 \cdot N \cdot M)_{cmac}$	да	да
Оценка и сглаживание периодограммы входного сигнала	$M_{add+mac}$	да	да
Оценка энергии шума (MCRA)	$M_{add+mac+cmp}$	да	да
Правило логарифмов спектральных амплитуд (LSA)	$M_{div+cmp} + M_{add+mpy+cmp}$	да	да
Расчет периодограммы «очищенной» речи	$M_{add+mpy+div} + M_{expint}$	частично	да
Оценка порогов маскирования	$(10N)_{mac} + 18_{pow+div+sub}$	частично	да
Перцептуальное взвешивание	M_{mpy}	да	да
WDFT-синтез	$(2N \cdot M)_{cmac}$	да	да
Формирование фрейма очищенного сигнала	$N_{add+mov} + 0.5N_{mov}$	да	да
Усреднение порогов маскирования	$M_{add+mpy}$	Да	да
Расчет субполосной перцептуальной энтропии	$M_{add} + 0.5M_{mpy+div} + 1_{sqrt+log2}$	Частично	да

Анализ табл. 18.2, табл. 18.3 показывает, что наиболее емкими в вычислительном плане являются такие операции как WDFT-анализ и WDFT-синтез в препроцессоре редактирования шума, а также оценка частоты основного тона и многоуровневое векторное квантование субполосного сигнала возбуждения в кодере.

Особенностью алгоритмов, используемых в комбинированной системе компрессии речевого сигнала, является то, что они базируются на однотипных и повторяющихся операциях: сложение, вычитание, умножение, умножение с аккумулярованием, масштабирование и т.п. Кроме того, практически на всех этапах существует возможность распараллеливания вычислений в несколько потоков, что при соответствующей аппаратной поддержке может значительно

¹ Нижние индексы в столбце «Вычислительная сложность, операций» означают одноименные операции

сократить время обработки. Поэтому наиболее целесообразно для аппаратной реализации системы компрессии широкополосного речевого сигнала использовать микропроцессор с возможностью аппаратного распараллеливания вычислений, многоступенчатым конвейером и VLIW-архитектурой [16,21,22].

Табл. 18.3 – Вычислительная сложность широкополосного компрессора/декомпрессора

Этап вычислений	Вычислительная сложность, операций	Использование нескольких исполнительных устройств, оптимальная загрузка конвейера	Загрузка двух операндов за такт
Перцептуальное взвешивание фрейма входного сигнала	$437N_{mac} + 4L_{mov} + 68L_{mac} + 4L_{sub}$	да	Да
Поиск ОТ для LTP-фильтра	$68D_{max}L_{mac}, D = 32 \dots 160$	частично	Да
Многоуровневое квантование сигнала возбуждения	$b(132L_{mac} + 4L_{sub} + 32L_{sum}) + 4(D_{max} + 16)_{mov}$	да	Да
Квантование параметров	$1245_{mac} + 32b_{mac} + 57_{cmp} + 45_{mov} + 128_{log2}$	частично	Да
Упаковка в пакеты для передачи	$150_{and} + 64b_{and}$	да	Да
Распаковка данных	$150_{and} + 64b_{and} + 128_{pow2}$	частично	Да
Деквантование параметров	178_{mov}	частично	Да
Реконфигурация кодера и синтез сигнала возбуждения	$32L_{add} + 32L_{mov}$	да	Да
Синтез реконструированного сигнала	$68L_{mac} + 4(D_{max} + 16)_{mov} + 4L_{mov}$	да	Да

18.2.2 Организация вычислительного процесса

Процесс кодирования поступающей речевой информации можно представить в виде последовательности следующих действий:

- 1) прием очередного фрейма речевого сигнала длительностью 20 мс (320 отсчетов при частоте дискретизации 16 кГц);

- 2) организация буфера анализа с учетом 50% перекрытия для WDFT;
- 3) запуск системы редактирования шума (если она включена);
- 4) сложение выходных данных по методу перекрытия с накоплением;
- 5) запуск кодера;
- 6) упаковка данных;
- 7) передача в ПЛИС для последующей шифрации и посылки в линию связи через модем.

Процесс декодирования включает в себя следующие действия:

- 1) прием пакета данных из ПЛИС;
- 2) распаковка данных;
- 3) запуск декодера;
- 4) выдача очередного синтезированного фрейма сигнала через EDMA в аудиокодек.

В общем случае процедуры кодирования и декодирования выполняются последовательно, при этом необходимо уложиться в жестко фиксированный бюджет времени между аппаратными прерываниями от EDMA, поступающими с интервалом 20 мс. Обобщенная диаграмма времени исполнения процедур комбинированной системы кодирования широкополосной речевой информации представлена на рис. 18.11.

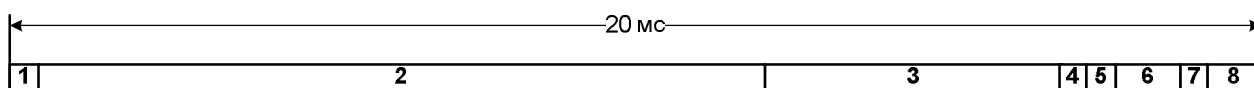


Рис. 18.11 – Временные затраты на кодирование и декодирование одного фрейма сигнала: 1– обработка прерывания от принимающего канала EDMA, 2–система редактирования шума (включена в программное прерывание кодера, вычислительный ресурс условно показан отдельно), 3–программное прерывание кодера, 4–аппаратное прерывание от ПЛИС (запрос пакета данных), 5–аппаратное прерывание от ПЛИС (прием пакета данных), 6–программное прерывание декодера, 7–аппаратное прерывание от передающего канала EDMA, 8–планировщик задач

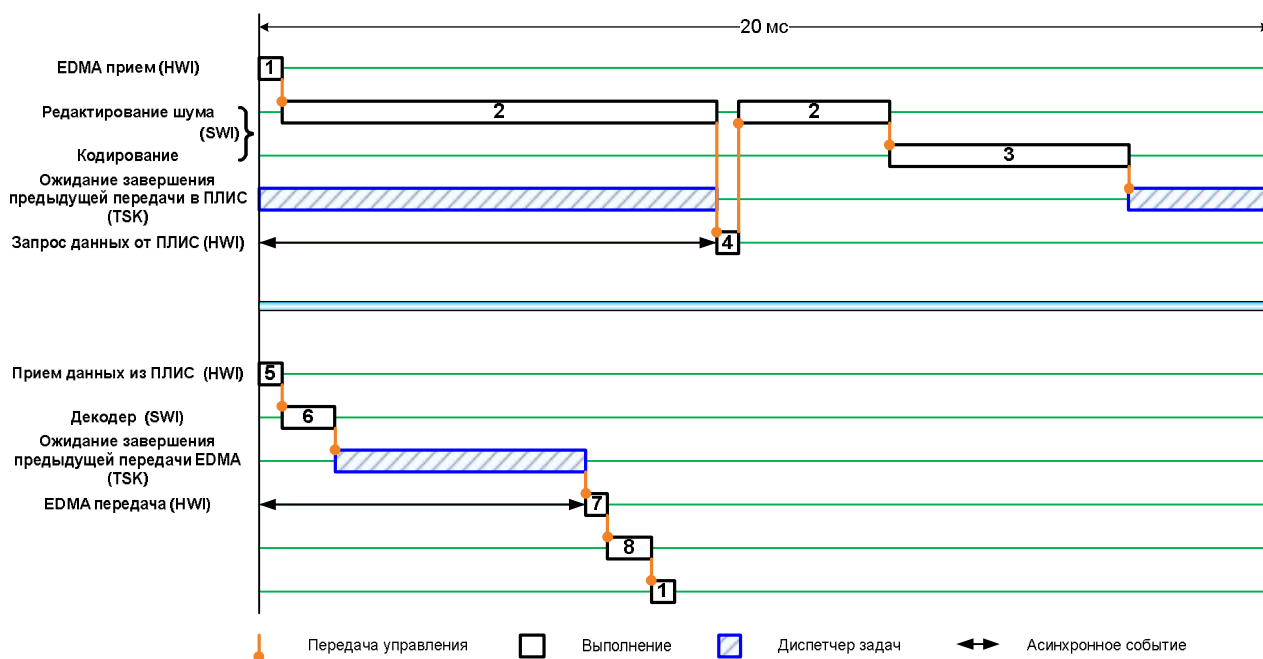


Рис. 18.12 – Организация вычислительного процесса во времени

Обмен пакетами через канал связи является асинхронным по отношению к обмену данными с аудиокодеком. То есть, прерывания от EDMA по приему и передаче аудиоданных могут приходиться в любой момент времени по отношению к прерываниям от ПЛИС (по приему и передаче пакетов данных). Поэтому используются дополнительные фоновые задачи для обеспечения синхронизации обмена пакетами с ПЛИС и фреймами сигнала с EDMA (аудиокодеком). Пример возможного расположения во времени аппаратных и программных прерываний представлен на рис. 18.12.

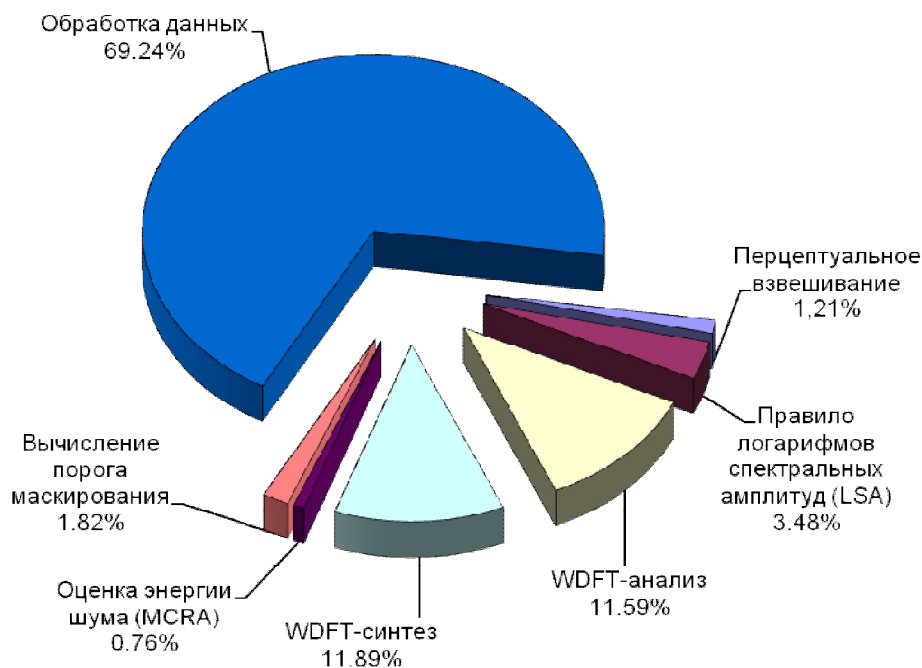


Рис. 18.13 – Распределение вычислительной сложности среди алгоритмов системы редактирования шума

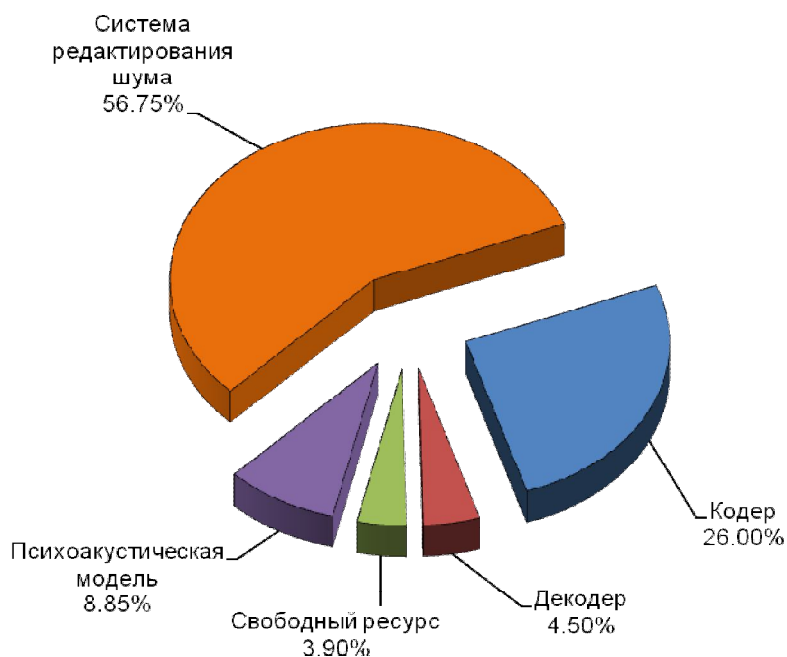


Рис. 18.14 – Вычислительная сложность комбинированной системы кодирования широкополосных речевых сигналов

Более детальное распределение вычислительного ресурса между составляющими алгоритмами препроцессора редактирования шума представлено на рис. 18.13, а общее распределение вычислительной сложности комбинированной системы кодирования широкополосного речевого сигнала – на рис. 18.14.

Оценка объема требуемой памяти, получена на основании карты, сгенерированной линковщиком и утилитой формирования кода, для записи во флэш-память (см. табл. 18.4).

Табл. 18.4 – Оценка требуемого объема памяти для комбинированной системы кодирования широкополосных речевых сигналов

Тип памяти	Использовано, Кбайт			Доступно, Кбайт
	Память программ	Память данных	Всего	
<i>IRAM</i>	178	10	188	192
<i>SDRAM</i>	67	3156	3223	8192
<i>EEPROM</i>	–	–	1895	2048

18.2.3 Оценки качества системы компрессии

Для оценки были записаны голоса для пяти дикторов (три мужских и два женских). Полученные оценки описываются отношениями сигнал/шум (*SNR*), шум/порог маскирования (*NMR*) [23] и параметрами искажения спектра барков, вычисленными по стандартному (*BSD*) и модифицированному (*MBSD*) [24] алгоритмам соответственно. Полученные результаты сведены в табл. 18.5.

Табл. 18.5 – Объективные оценки качества восстановленного в декодере речевого сигнала

Отношение сигнал/шум (<i>SNR</i>), дБ	Отношение шум/порог маскирования (<i>NMR</i>), дБ	Оценка искажений спектра барков (<i>BSD</i>)	Модифицированная оценка искажений спектра барков (<i>MBSD</i>)
11.59	-6.76	0.05	0.02

Используемые в настоящее время методы субъективных экспертных оценок качества речи [25] стандартизированы и регламентируются в следующих стандартах: ГОСТ 51061-97 «Параметры качества речи и методы ее измерения» (Low bit rate speech transmission systems. Speech quality characteristics and their evaluation), ГОСТ Р 50840-95 (СТБ ГОСТ Р 50840-2000) «Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости» (Speech transmission over varies communication channels. Techniques for measurements of speech quality, intelligibility and voice identification), международные стандарты ISO/TR4870, IEC 268-16, ANSI S3.2-1989 «Method for measurement the Intelligibility of Speech over Communication Systems».

Несмотря на то, что сейчас существует большое количество объективных критериев оценки качества речи и широко используется специальное программно-аппаратное обеспече-

ние для их реализации, оценки разборчивости речи с помощью квалифицированных экспертов по-прежнему считаются достаточно достоверными и все новые объективные критерии сравниваются с ними [25].

В результате тестов были получены следующие результаты [26] – табл. 18.6.

Табл. 18.6 – Субъективные оценки качества звучания реконструированной речи

Слоговая разборчивость речи, %	Субъективное качество речи, баллы	Особенности восстановленной в декодере речи	Узнаваемость голоса диктора, %
98	4.2	Хорошая естественность и узнаваемость, иногда присутствует подзванивание	95

В работах [16] и [18] был проведен сравнительный анализ качества разработанной модели вокодера со стандартизированными моделями подобного класса (ITU-T G.722 [27,28] и MPEG-1 Layer III [29]). Результаты сравнения представлены на рис. 18.15 и рис. 18.16, соответственно.

Сравнительный анализ разработанного компрессора с кодеком MPEG1 Layer III (см. рис. 18.15) показывает, что предлагаемый кодер с динамически реконфигурируемой структурой банка кодовых книг обеспечивает качество реконструированной речи, сравнимое с mp3-кодеком на 24 кбит/с, в то же время кодер с предельной структурой банка кодовых книг имеет качество сравнимое с mp3-кодеком на 32 кбит/с.

Анализ оценок качества, представленных на рис. 18.16, показывает, что предлагаемый кодер с предельной и реконфигурируемой структурой обеспечивает качество выше, чем кодек ITU-T G.722 [27] на 64 кбит/с при меньшей скорости потока данных, что свидетельствует об эффективности разработанного кодера.

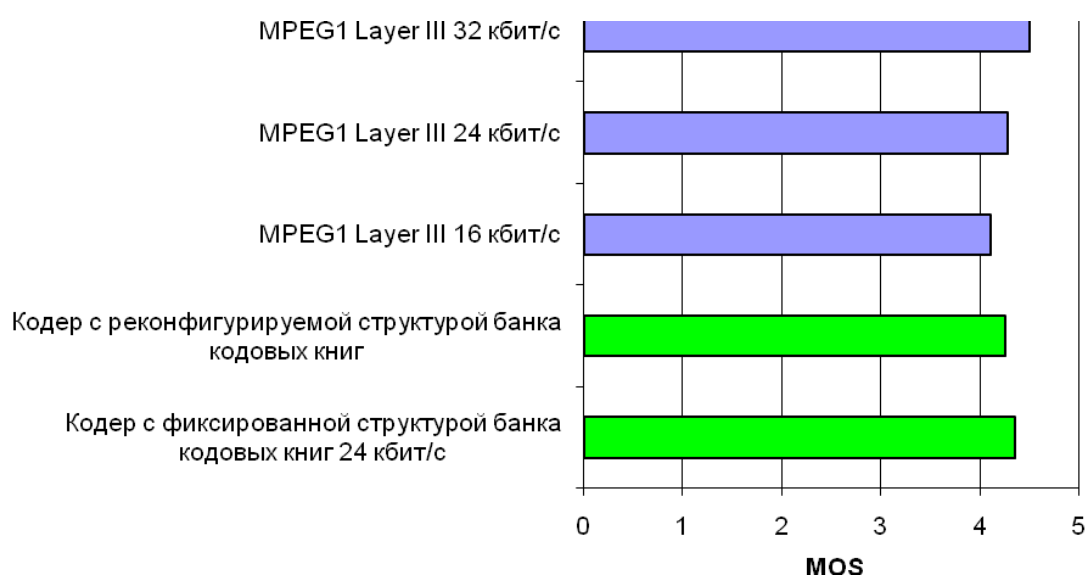


Рис. 18.15 – Сравнение качества по процедуре MOS предлагаемого кодера и MPEG1 Layer III кодека

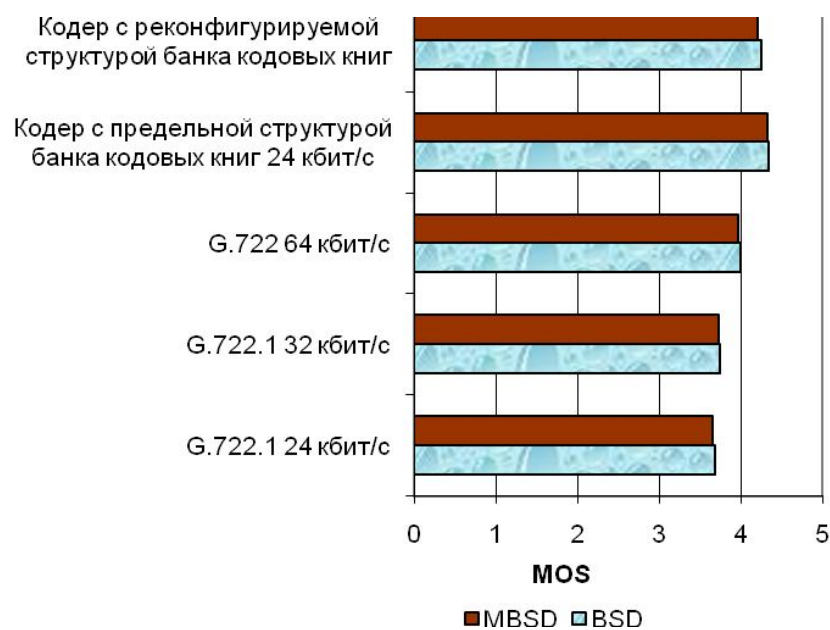


Рис. 18.16 – Сравнение с кодеком ITU-T G.722

18.2.4 Объективная и субъективная оценка качества препроцессора редактирования шума

Искажение речевого сигнала было измерено, используя сегментное отношение сигнал к шуму (SNR_{seg}), где шум был интерпретирован как разница между оригинальным и очищенным речевым сигналом [30]. При этом, чем выше значение SNR_{seg} , тем слабее искажение речи. Рис. 18.17а показывает результаты, полученные для различных типов шума и его уровней.

Так как мера SNR_{seg} основана на простом сравнении энергии в речевых фреймах, она не говорит ничего о подобности сравниваемого спектра. По этой причине была использована мера кепстрального расстояния (от англ. – «*Cepstral Distance*» – CD) [30]. Произведена оценка расстояния между чистым и очищенным спектром сигнала. В этом случае большая величина меры CD отражает более сильное искажение речи. Результаты представлены на рис. 18.17б.

Степень подавления шума была измерена при помощи коэффициента ослабления шума NA [30], который определяется как среднее отношения мощности шума на входе системы редактирования шума к мощности остаточного шума на выходе системы (см. рис. 18.17в).

Основной недостаток инструментальных измерений заключается в том, что они имеют относительно малую степень корреляции с восприятием речи. Это обусловлено тем, что данные меры искажений часто используют только статистические модели для расчета чистой разницы между сигналами вместо использования свойств системы восприятия человека. По этой причине, для определения воспринимаемых различий между оригинальным и очищенным сигналом необходимы перцептуальные меры. В данном случае была использована мера $MBSD$ [24], так как эта оценка наиболее коррелирована с субъективной оценкой качества. Результаты представлены на рис. 18.17г.

Объективные оценки качества системы редактирования шума показывают, что предлагаемый подход обеспечивает достаточно высокую степень подавления шума при сохранении высокого перцептуального качества. При этом система одинаково хорошо работает с шумами

различной природы, о чем свидетельствуют оценки ослабления шума и $MBSD$, представленные на рис. 18.17в и рис. 18.17г, соответственно. Анализ рис. 18.17в показывает, что препроцессор повышения качества речи обеспечивает максимальную степень подавления для «белого» шума. Наименьшее ослабление характерно для «коричневого» шума. Приблизительно одинаковые показатели ослабления характерны для следующих шумов: «розовый» шум, шум автомобиля и гул в помещении.

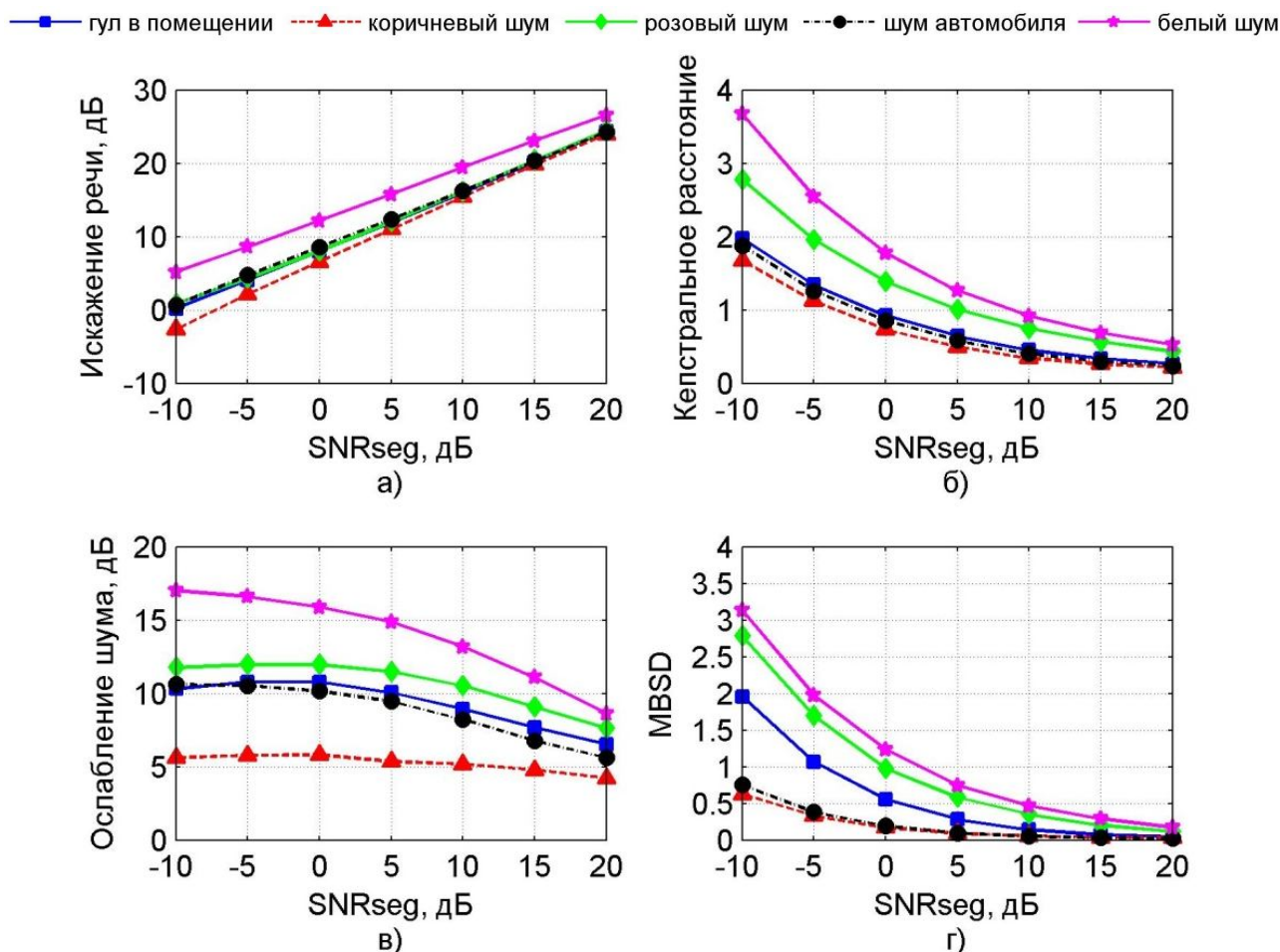


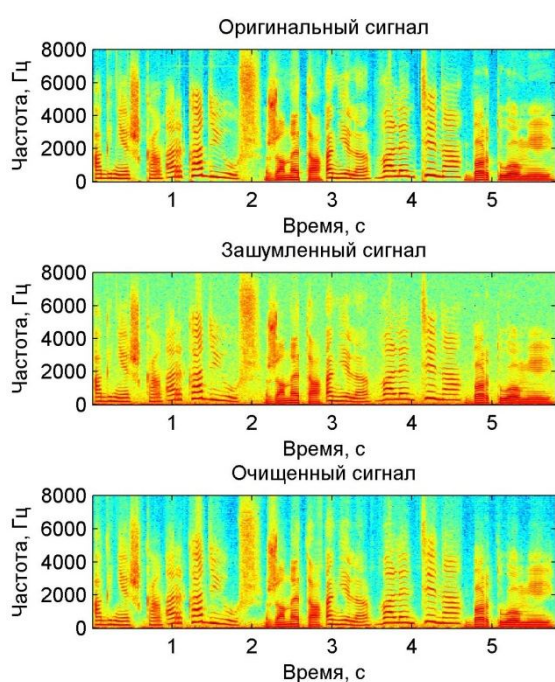
Рис. 18.17 – Объективные оценки качества системы редактирования шума

Для проведения анализа структуры остаточного шума различной природы были построены спектрограммы (см. рис. 18.18). С целью проведения оценки были выбраны следующие типы шума: «коричневый», «белый», шум автомобиля и гул помещения. На один и тот же оригинальный речевой файл был наложен шум соответствующего типа с энергией обеспечивающей соотношение $SNR_{seg} = 0$ дБ.

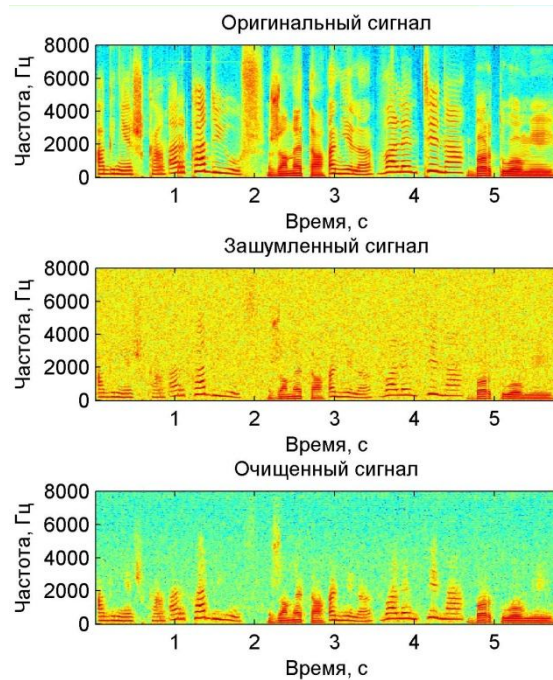
Анализ спектрограмм показывает, что система редактирования шума обеспечивает малое искажение речи и высокую степень подавления шума. Это обусловлено более точной оценкой СПМ в критических частотных полосах. Во всех случаях «музыкальный тон» не наблюдается благодаря относительно высокому заранее определенному уровню остаточного шума.

Таким образом, интегрированная система подавления шума расширяет количество мобильных мультимедийных приложений, в которых может работать комбинированная система. Это обеспечивается гибкостью настройки параметров препроцессора повышения качества

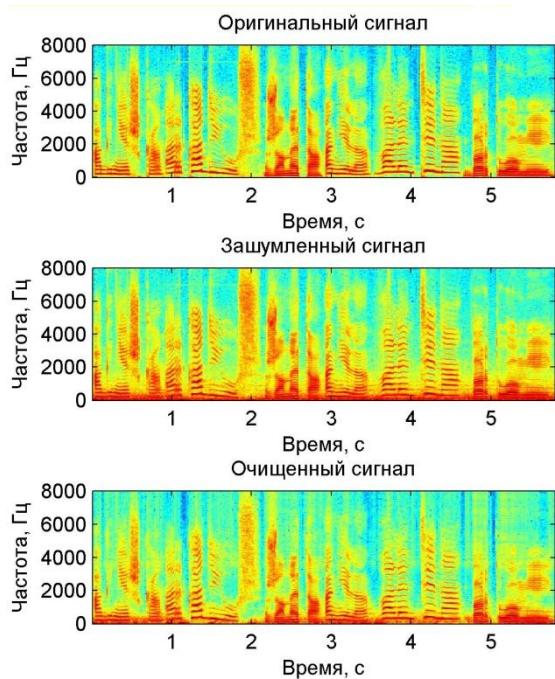
речи, что позволяет подстраивать систему к конкретной окружающей обстановке, в которой предполагается использовать систему компрессии речевых сигналов. При этом возможно задание степени максимального подавления шума, уровня остаточного шума и т.п.



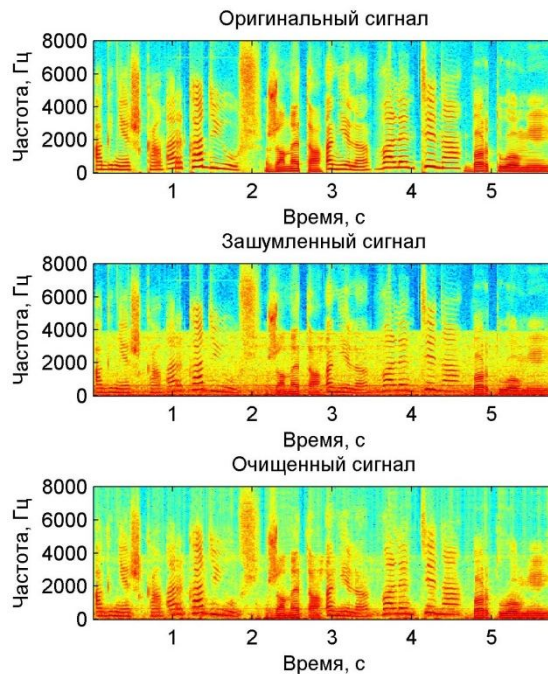
а) «коричневый» шум



б) «белый» шум



в) шум автомобиля



г) гул помещения

Рис. 18.18 – Результаты работы препроцессора повышения качества речи

18.2.5 Эффективность комбинированной системы кодирования звуковых сигналов

Для осуществления общего анализа качества комбинированной системы кодирования речевых сигналов последняя была протестирована в условиях окружающих шумов различной природы. Проведена оценка изменения перцептуального качества зашумленной речи. В качестве исходного речевого материала использовалась база данных CORPORA [31]. Осуществлено сравнение оценок BSD и $MBSD$ для реконструированной зашумленной речи и реконструированной очищенной речи обработанной препроцессором повышения качества входного речевого сигнала. Сравнительный анализ усредненных оценок BSD и $MBSD$ для различных шумов представлен на рис. 18.19. Изменение усредненного отношения NMR представлено на рис. 18.20. Улучшение оценок BSD и $MBSD$ проиллюстрировано на рис. 18.21.

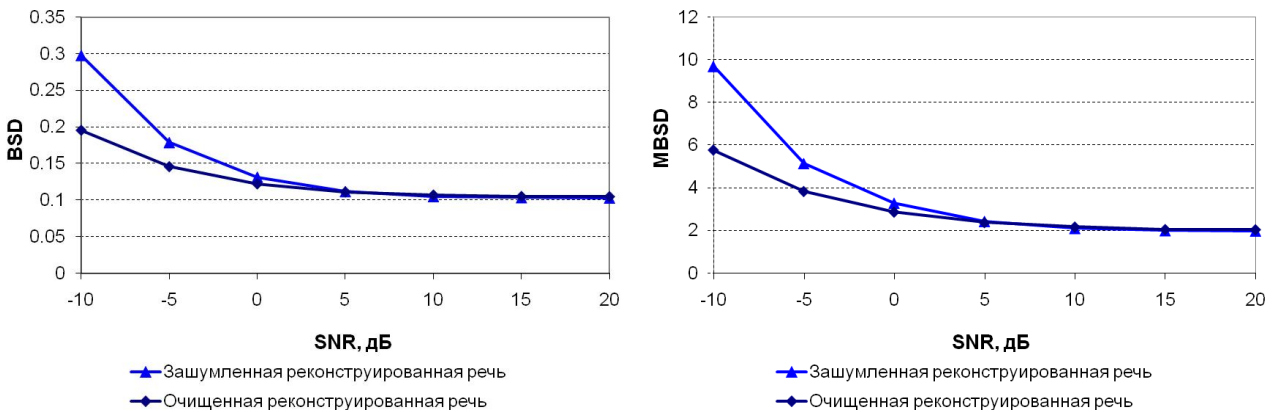


Рис. 18.19 – Влияние системы редактирования шума на изменение оценок BSD и $MBSD$ реконструированной речи

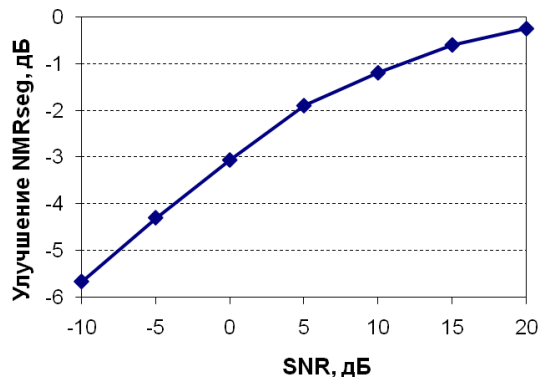


Рис. 18.20 – Изменение усредненной оценки NMR_{seg} для шумов различной природы с разными уровнями

На основании анализа и оценка качества работы комбинированной системы кодирования широкополосной речи в условиях окружающих шумов разного уровня и различной природы была определена область эффективной работы препроцессора повышения качества зашумленной входной речи (см. рис. 18.22).

Анализ оценок NMR_{seg} , BSD , $MBSD$ показывает, что система редактирования в составе комбинированной системы кодирования широкополосной речевой информации обеспечивает улучшение перцептуального качества реконструированной речи в условиях окружающих шумов

с диапазоном SNR от -10 дБ до +6 дБ. Для SNR более 6 дБ, система редактирования шума не влияет на качество реконструированной речи, что обусловлено мультиполосной моделью кодирования в широкополосном коде, которая обеспечивает кодирование вместе с голосом диктора и окружающей обстановки. То есть данные шумы являются комфортными и кодек достаточно хорошо справляется с их кодированием.

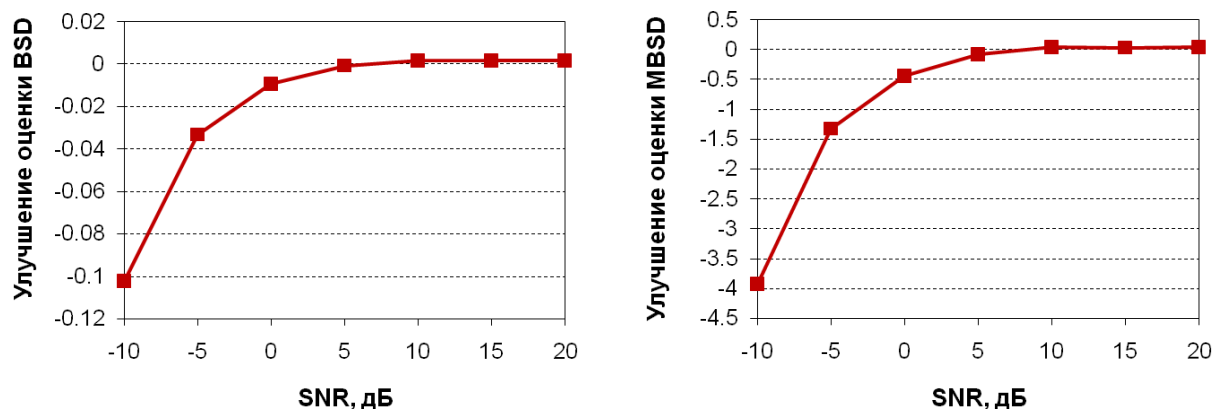


Рис. 18.21 – Влияние препроцессора повышения качества речи на улучшение оценок BSD и $MBSD$

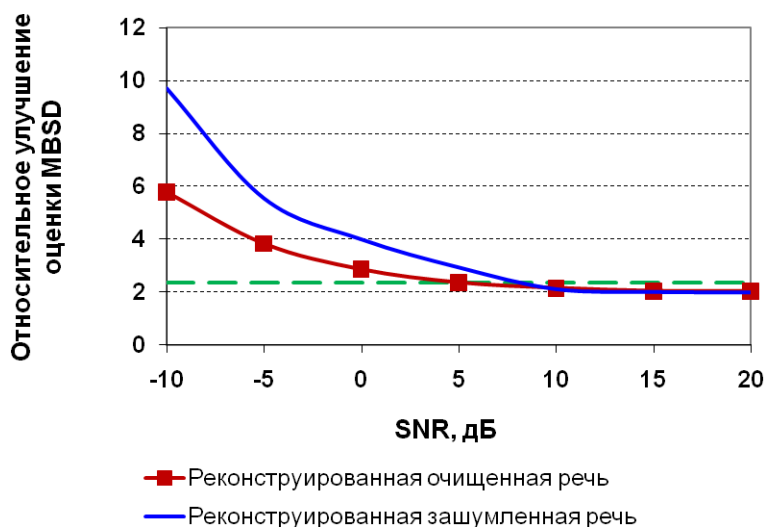


Рис. 18.22 – Область эффективной работы системы редактирования шума в условиях окружающих шумов разного уровня и различной природы

В целом, экспериментальные исследования показывают достаточно высокие показатели качества комбинированной системы кодирования широкополосных речевых сигналов, что свидетельствует о выборе правильного подхода для проектирования перцептуальных систем подобного класса.

18.3 Реализация кодера звука на базе ПДВП с динамической трансформацией частотно-временного плана на вычислительной платформе в реальном времени

18.3.1 Структура ПДВП-кодера звука

На рис. 18.23 представлена структура перцептуального кодера звукового сигнала, ядром которой является блок «WPD», выполняющий реализацию адаптивного пакета дискретного вейвлет преобразования (ПДВП). В темпе обработки сигналов для каждого текущего дерева E_i вычисляются пороги маскирования $T_{l,n}$, в соответствии с [32], значения перцептуальной энтропии $PE_{l,n}$, а также энтропии H_{E_i} структуры дерева ПДВП E_i . На основании данной информации в блоке «Split» рассчитываются параметры реконфигурации дерева ПДВП $r_{l,n}$ [33]. Данный процесс осуществляется поступательно, без возвратов на меньшие масштабные уровни преобразования. Следовательно, вычислительный процесс представляет собой структуру поточного процессора [34].

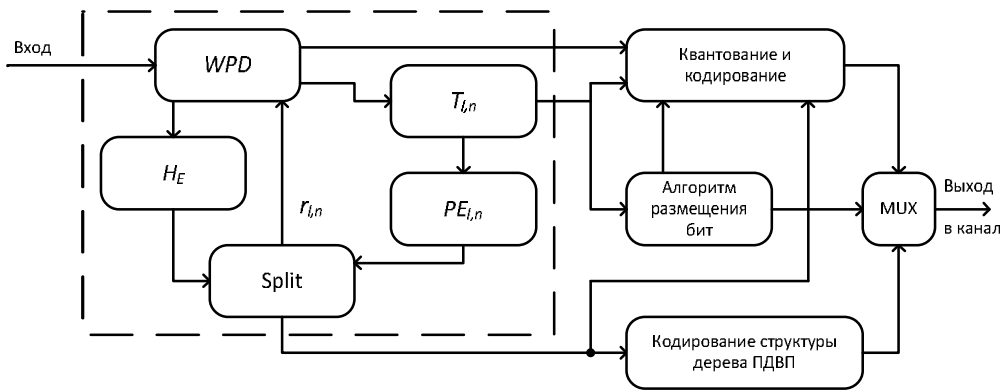


Рис. 18.23 – Структура ПДВП кодера звука

Обработка звукового сигнала в ПДВП кодере выполняется кадрами и ввиду того, что преобразование осуществляется над каждым последующим кадром в новой структуре дерева преобразования, то для устранения фазовых разрывов, соседние кадры сигнала перекрываются на $1/8$ своей длины. Следующим этапом работы кодера является квантование и кодирование вейвлет коэффициентов $X_{l,n,k}: (l,n) \in E, k \in \mathbb{Z}$, результирующего субоптимального дерева ПДВП E . Управление данным процессом осуществляется в блоке «Алгоритм размещения бит». Кодирование структуры дерева ПДВП E выполняется в блоке «Кодирования структуры дерева ПДВП». Принимая во внимание факт, что рост дерева осуществляется поступательно и изменчивость сигнала во времени более инерционна, чем время обработки, то кодируются только изменения структуры дерева от фрейма к фрейму.

Структура ПДВП декодера звука показана на рис. 18.24. Работа декодера выполняется в следующем порядке: входная информация разделяется на два потока данных в блоке «DEMUX»: один содержит кодированные вейвлет коэффициенты, а второй – код структуры дерева ПДВП; формирование структуры дерева ПДВП происходит в блоке «Формирование структуры дерева ПДВП» и осуществляется путём суммирования кодовой последовательности, описывающей изменение структуры дерева ПДВП, с текущей кодовой последовательностью струк-

туры дерева ПДВП. В случае передачи первого фрейма, структура дерева ПДВП кодируется полностью в кодере. Реконструкция звукового сигнала выполняется в блоке « WPD^{-1} » на основе обратного адаптивного ПДВП, над декодированными и восстановленными вейвлет коэффициентами (блок «Декодирования и восстановления») в соответствии со структурой дерева ПДВП. Синтезированные кадры реконструированного сигнала умножаются на модифицированное треугольное окно и складываются для формирования выходной последовательности $\hat{x}(n)$.

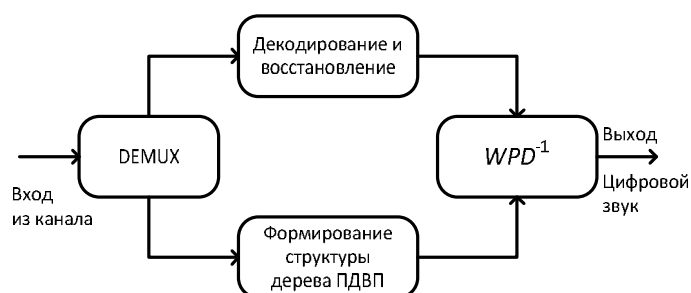


Рис. 18.24 – Структура ПДВП декодера звука

Структура вычислительной платформы кодера звука на основе подхода динамической трансформации алгоритма [33] включает управляющий процессор на базе ЦПОС и периферийный процессор – ПДВП процессор с реконфигурируемой архитектурой (рис. 18.2). На ЦПОС возложены следующие задачи: обработка вейвлет коэффициентов $X_{l,n,k}$ в частотных полосах (l, n) , соответствующих текущей структуре дерева E_i ; вычисление психоакустической модели дерева $(l, n) \in E_i: T_{l,n}$ и $PE_{l,n}$; получение вектора реконфигурации ПДВП процессора $r_{l,n}$, $(l, n) \in E_i$; формирование и разборка данных передачи $[code(X_{l,n,k}, (l, n) \in E_i, k \in \mathbb{Z})]$ и приема $\hat{X}_{l,n,k}(code)$, $(l, n) \in E_i, k \in \mathbb{Z}$, если используется дуплексный режим работы кодера звука. ПДВП процессор реализуется на параллельно-поточной архитектуре с динамической реконфигурацией. Выбор данной архитектуры обусловлен алгоритмом роста дерева ПДВП. Процессор работает в дуплексном режиме: конвейер кодера $X_{l,n,k} = WPD_{l,n}(x_j(t))$, $(l, n) \in E_i$, $x_j(t)$ – j -й фрейм звукового сигнала и конвейер декодера $\hat{x}_j(t) = WPD_{l,n}^{-1}(\hat{X}_{l,n,k})$, $(l, n) \in E_i$, $\hat{x}_j(t)$ – j -й фрейм реконструированного звукового сигнала. Длина конвейера параллельно-поточного процессора определяется предельной структурой дерева ПДВП ($CB - WPD$). Анализ ёмкостной сложности реализации базовой операции ПДВП (пара КИХ фильтров нижних и верхних частот с децимацией), модулей коммутации и буферизации, требуемый аппаратный ресурс ПДВП процессора следующий: объём КЛБ для построения 8-ми модулей базовой операции ПДВП на МАС архитектуре – 2048, и для схем буферизации и переключения – 504. Расчетное число КЛБ с учётом 20% надбавки, обусловленной накладными расходами при реализации на конкретной ПЛИС, равно 3062 КЛБ.

18.3.2 Организация вычислительного процесса

Процесс кодирования звукового сигнала можно представить в виде последовательности следующих действий: прием очередного кадра сигнала длительностью 23 мс (1024 отсчета при частоте дискретизации 44.1 кГц); передача в ПЛИС вектора реконфигурации $r_{l,n}$; анализ входного сигнала; расчет параметров психоакустической модели на ЦПОС; формирование век-

тора реконфигурации ПДВП процессора задачи анализа; кодирование вейвлет коэффициентов на управляющем процессоре; формирование пакетов передачи кодированной информации через цифровой порт.

Следует отметить, что работа кодера звука в реальном масштабе времени обуславливает достаточно жесткую синхронизацию вычислительных процессов управляющего процессора и ПДВП процессора. Так перестройка ПДВП процессора осуществляется на основе формирования вектора реконфигурации $r_{l,n}$ согласно алгоритму динамической трансформации дерева ПДВП. При этом для фреймов, обрабатываемых согласно текущей структуре дерева $E_{m,i}$ на ступенях ПДВП процессора m , управляющий процессор осуществляет выполнение процедур расчета порогов маскирования, оценки перцептуальной энтропии и информативности структуры дерева ПДВП. Распределение вычислительного ресурса между процедурами в управляющем процессоре показано на рис. 18.25. Время выполнения процедуры 1 (алгоритм вычисления порогов маскирования (рис. 18.26)) и процедуры 2 (оценка перцептуальной энтропии (рис. 18.26)) не постоянно, оно зависит от числа ступеней m конвейера ПДВП процессора (кодер), задействованных в обработке текущего фрейма входного сигнала (см рис. 18.27). Этой же причиной, обусловлено непериодическое выполнение процедур 4 и 5 (квантование и кодирование вейвлет коэффициентов $X_{l,n,k}$).

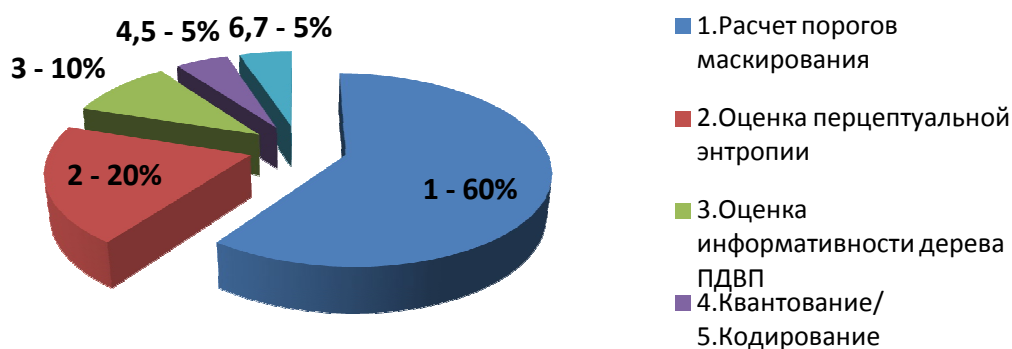


Рис. 18.25 – Распределение вычислительного ресурса между процедурами в управляющем процессоре

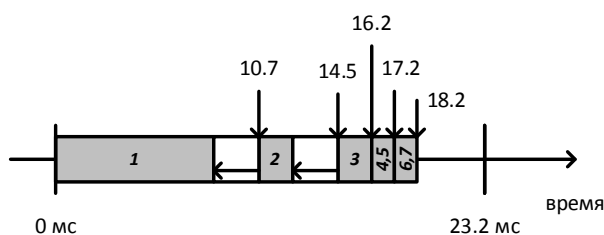


Рис. 18.26 – Время выполнения процедур в управляющем процессоре

В режиме декодирования управляющий процессор декодирует и восстанавливает фреймы из принимаемой кодовой последовательности, выполняя процедуры 6,7 (декодирование и восстановление), загружает восстановленный фрейм i на соответствующую ступень m конвейера ПДВП процессора (декодера). Переход фрейма i на следующую ступень процессора выполняется тогда, когда управляющий процессор примет следующий фрейм $i + 1$, который

необходимо будет загрузить уже на задействованную ступень конвейера ПДВП процессора (декодера) (см. рис. 18.28). Для согласования работ, выполняемых на каждой ступени конвейера процессора, необходимо ввести задержку, кратную времени обработки одного фрейма на одной ступени, тем самым будет обеспечиваться ритмичная работа параллельно-поточной структуры ПДВП процессора (декодер).



Рис. 18.27 – Схема анализа входного сигнала в ПДВП процессоре

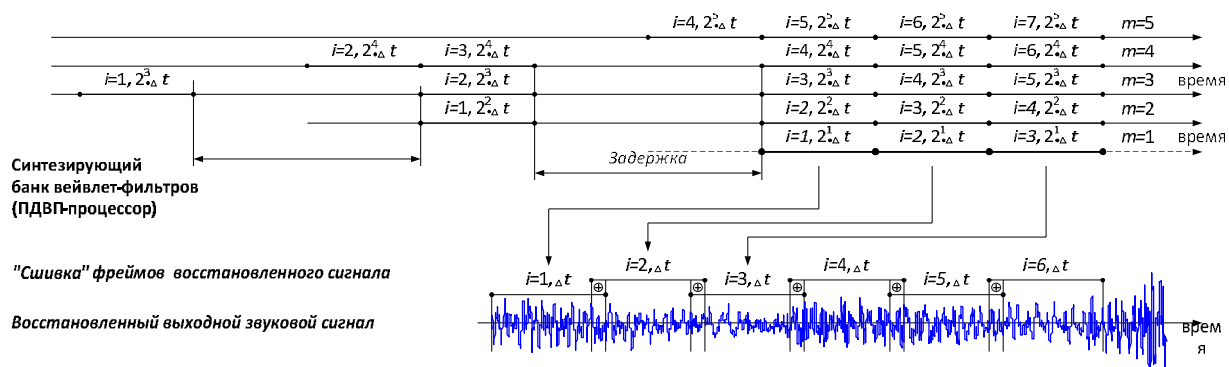


Рис. 18.28 – Синтез выходного сигнала в ПДВП процессоре

18.3.3 Оценка эффективности кодера

Ниже представлены результаты исследований данного субполосного перцептуального ПДВП кодера звука. Для разных сигналов (шумовых, тональных, ритмичных) оценивалась возможная минимальная скорость передачи (кбит/с), максимальная степень компрессии, среднее количество бит на отсчет входного звукового сигнала, а также определялись объективные оценки качества: NMR_{total} , NMR_{SEG} – общее и сегментное соотношение шума к порогу маскирования, показывающее границу порога восприятия артефакторов на слух, SB - ширина полосы частот реконструированного сигнала, MF - процент искаженных фреймов реконструированного сигнала (см. рис. 18.29) [23].

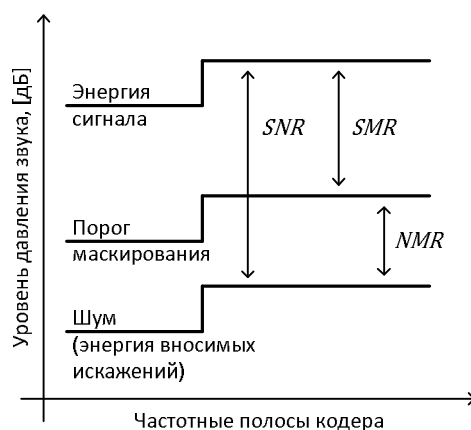


Рис. 18.29 – Объективные оценки качества сигнала

Табл. 18.7 – Оценки перцептуального ПДВП кодера звука и MPEG-1 уровень III

Тестовый материал $f_s = 44.1$ кГц, 16 бит ИКМ	Тип кодера	Мин. скорость передачи, [кбит/с]	Макс. степень компрессии	Среднее кол., [бит/отсчет]	Ширина частотной полосы SB , [кГц]	NMR_{total} , [дБ]	NMR_{SEG} , [дБ]	MF , % искаженных фреймов
Boney-M 15с.	ПДВП-кодер	45	15.68	1.02	18.49	-8.297	-0.8699	27.04
	MPEG-1 ур. III	64	11.02	1.45	15.13	4.430	0.4076	99.85
Classic 10с.	ПДВП-кодер	42	16.8	0.95	17.62	-6.495	-0.7987	43.45
	MPEG-1 ур. III	64	11.02	1.45	12.58	3.284	0.3145	99.78
ABBA "Take a chance on me" 2с.	ПДВП-кодер	36	19.7	0.81	18.58	-4.037	-0.5844	66.25
	MPEG-1 ур. III	64	11.02	1.45	15.39	3.720	0.3312	98.70

В табл. 18.7 приведены результаты экспериментов в сравнении с популярным кодером звука MPEG-1 уровень III [35], минимальная скорость передачи которого 64 кбит/с. Анализ данных табл. 18.7 показывает, что минимальная скорость передачи для ПДВП кодера звука меняется от 36 кбит/с для тональных, ритмичных звуковых сигналов до 45 кбит/с для шумовых сигналов. При этом, оценка NMR_{total} меняется от -4,037 дБ до -8,857 дБ, т.е. граница порога восприятия минимум в два раза ниже порога маскирования. Для сравнения у MPEG-1 данная оценка позитивная, т.е. энергия воспринимаемых шумов не менее чем в два раза больше порога маскирования и, соответственно, практически каждый фрейм реконструированного сигнала содержит искажения и, $MF_{max} \cong 99,78\%$. Это обусловлено скорее тем, что ширина частотной полосы SB реконструированного сигнала достаточно мала: $SB_{min} = 11,89$ кГц, против 22,05 кГц в оригинальном сигнале (см. рис. 18.30 - Рис. 18.32).

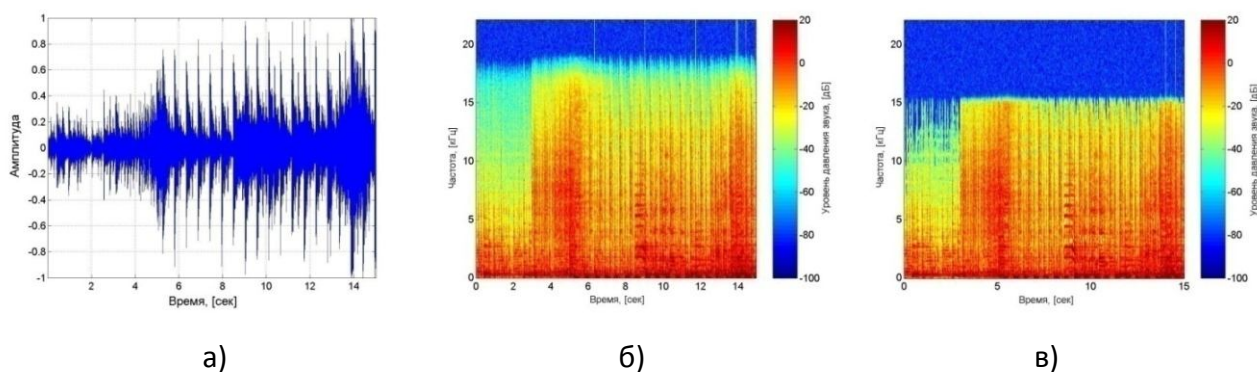


Рис. 18.30 – Звуковой сигнал «BoneyM-2000»: а) оригинальный сигнал; б) спектрограмма реконструированного сигнала (ПДВП-кодер); в) спектрограмма реконструированного сигнала (MPEG-1 уровень III)

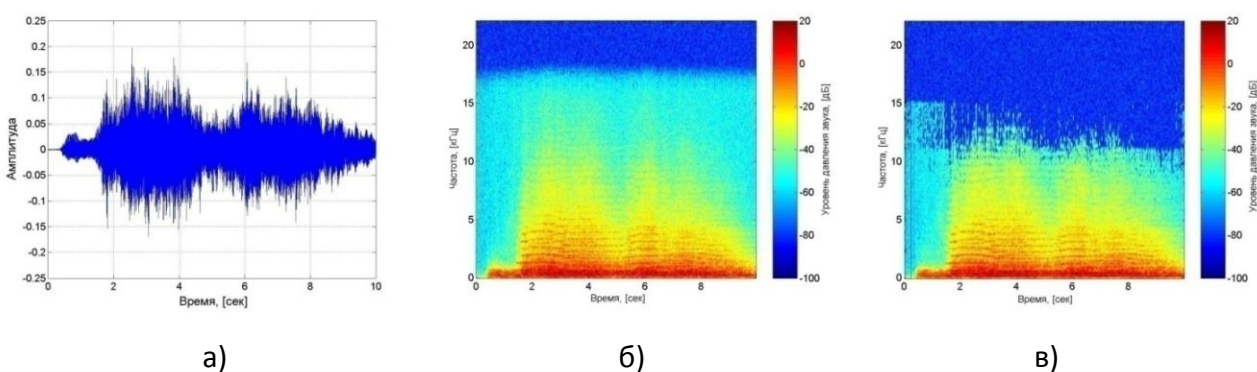


Рис. 18.31 – Звуковой сигнал «Classic»: а) оригинальный сигнал; б) спектрограмма реконструированного сигнала (ПДВП кодер); в) спектрограмма реконструированного сигнала (MPEG-1 уровень III)

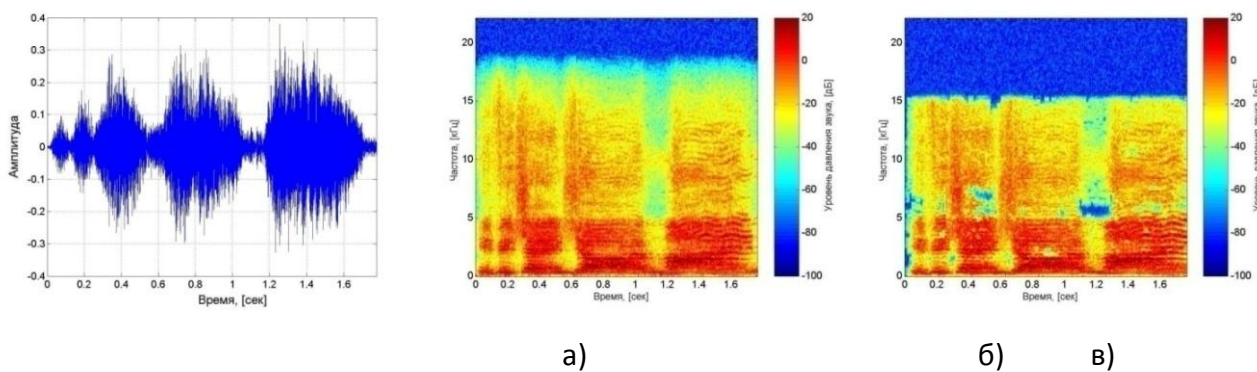


Рис. 18.32 – Звуковой сигнал «ABBA “Take a chance on me”»: а) оригинальный сигнал; б) спектрограмма реконструированного сигнала (ПДВП кодер); в) спектрограмма реконструированного сигнала (MPEG-1 уровень III)

Таким образом, объективные оценки, полученные при экспериментальных исследованиях перцептуального ПДВП кодера звуковых сигналов, показывают, что реконструированный сигнал имеет достаточно хорошее качество, соответствующее требованиям стандарта ITU-R REAQ [36] при высокой степени компрессии в 15 и более раз. Причём, для NMR_{total} меньше -10 дБ считается, что восстановленный сигнал не содержит слышимых артефактов. Для данного

субполосного перцептуального ПДВП кодера звука данный показатель ≈ -9 дБ, но для схемы измерения NMR_{total} , главным достоинством которой, а также оценок, MF и SB , является то, что в отличие от [23] они вычисляются в вейвлет области и при тестировании кодеров звука на базе адаптивных ПДВП дают минимальную погрешность вследствие отсутствия взаимных частотно-вейвлетных преобразований [37,38].

18.4 Список использованных источников

1. TMS320C6713B Floating-Point Digital Signal Processor, SPRS294B, Texas Instruments, Dallas, TX, June, 2006, 152 p.
2. Livshitz M., Petrovsky Al., Stankevich A., Kachinsky M., Petrovsky A., Universal Embedded Reconfigurable Hardware Platform for Multimedia Applications in Real-Time, The International Journal of Computing, 2008, Vol. 7, Issue 3. pp. 38-46.
3. Spartan–3 FPGA Family: Complete Data Sheet (DS099) [Electronic resource] / Xilinx Inc., May, 2007, Mode of access: <http://direct.xilinx.com/bvdocs/publications/ds099.pdf>, Date of access: 25.05.2007.
4. Зотов В.Ю., Проектирование цифровых устройств на основе ПЛИС фирмы Xilinx в САПР WebPack ISE, М.: Горячая линия – Телеком, 2003, 624 с.
5. TLV320AIC23B Stereo Audio Codec, Data Manual, SLWS106H, Texas Instruments, Dallas, TX, February, 2004, 56 p.
6. Creating a Second-Level Bootloader for FLASH Bootloading on TMS320C6000 Platform with Code Composer Studio, SPRA999A1, Texas Instruments, Dallas, TX, May, 2006, 27 p.
7. TMS320C6000 DSP Software-Programmable Phase-Locked Loop (PLL) Controller Reference Guide, SPRU233C, Texas Instruments, Dallas, TX, August, 2004, 24 p.
8. TMS320C6000 DSP External Memory Interface (EMIF) Reference Guide, SPRU266E, Texas Instruments, Dallas, TX, February, 2006. 146 p.
9. Аппаратно-программная реализация широкополосного компрессора речевых сигналов с психоакустической мотивацией для применения в системах мультимедиа реального времени на базе ЦПОС TMS320C6X: отчет о НИР (заключительный), БГУИР; рук. темы А.А. Петровский, Минск, 2007, 22 с, № ГР 20063161.
10. Разработка широкополосного компрессора речевых сигналов с психоакустической мотивацией на базе мультиполосного CELP-алгоритма для низкоскоростных каналов связи: отчет о НИР (заключительный), БГУИР; рук. темы А.А. Петровский, Минск, 2006, 40с, № ГР 2005761.
11. TMS320C6000 DSP Enhanced Direct Memory Access (EDMA) Controller Reference Guide, SPRU234B, Texas Instruments, Dallas, TX, March, 2005, 269 p.

-
12. Parfieniuk M., Petrovsky A., Warped DFT as the basis for psychoacoustical model, The proc. of the IEEE International conference on Acoustic, Speech, Signal processing, ICASSP-2004, Montreal, 2004, pp. IV-185-188
 13. Petrovsky A., Parfieniuk M., Borowicz A., Petrovsky A., Auditory modeling via frequency warped transforms, Audio Engineering Society, Convention Paper 7449, presented at the 124th Convention, 17-20 May, Amsterdam 2008, The Netherlands, 15 pages.
 14. Petrovsky Alexander A., Parfieniuk M., Borowicz A., Warped DFT based perceptual noise reduction system, Signal processing: proceedings of 116th AES convention, Germany, Berlin, May 8–11, 2004, Berlin, Germany, 2004, Preprint № 6035, 16 pages.
 15. Gustafsson S., Jax P., Vary P., A novel psychoacoustically motivated audio enhancement algorithm preserving background noise characteristic, IEEE Int. Conf. on Acoustic, Speech, and Signal Processing ICASSP'98, Seattle, USA, 1998, Vol. 1, pp. 397–400.
 16. Livshitz M., Parfieniuk M., Petrovsky A.A., Real-Time Implementation of Combined Noise Reduction and Wideband Speech Coding Systems based on the DSP TMS320C6713, in "New Trends in Audio and Video", edited by Dobrucki A., Petrovsky A., Skarbek W., Politechnika Białostocka, Białystok, 2006, Vol. I, pp. 473–478.
 17. Лившиц М.З., Парфенюк М., Петровский А.А., Широкополосный CELP-кодер с мультиполосным возбуждением и многоуровневым векторным квантованием по кодовой книге с реконфигурируемой архитектурой, Цифровая обработка сигналов, №2, Москва, 2005, с. 20-35.
 18. Livshitz M., Petrovsky A.A., Perceptually Constrained Variable Bitrate Wideband Speech Coder, The Proc. of EUROCON, Serbia & Montenegro, Belgrade, November 22-24, 2005, pp. 1296-1299.
 19. TMS320C67x FastRTS Library Programmer's Reference [Electronic resource] / Texas Instruments, 2002, Mode of access: <http://focus.ti.com/lit/ug/spru100a/spru100a.pdf>, Date of access: 15.05.2007.
 20. Borowicz A., Parfieniuk M., Petrovsky A.A., An application of the warped discrete Fourier transform in the perceptual speech enhancement, Speech Communication, vol. 48, ELSEVIER, 2006, pp.1024-1036
 21. Fisher J.A., Faraboschi P., Young C., Embedded Computing: A VLIW Approach to Architecture, Compilers, and Tools, Elsevier, 2005, 709 p.
 22. Livshitz M., Petrovsky A.A., An Overcomplete WDF-based Perceptually Constrained Variable Bit Rate Wideband Speech Coder with Embedded Noise Reduction System, The Proc. of XI International Conference "SPEECH AND COMPUTER", St. Petersburg, Russia, June 25-29, 2006, pp. 343-348.

-
23. Brandenburg K., Sporer T., "NMR" and "masking flag": Evaluation of quality using perceptual criteria, Audio test and measurement: proceedings of 11th AES int. conf., Portland, USA, May 29–31, 1992, pp. 169–179.
 24. Yang W., Benbouchta M., Performance of the modified bark spectral distortion as an objective speech quality measure, Acoustic, Speech and Signal Processing: proceedings of IEEE int. conf. (ICASSP'98), USA, Seattle, May 12–15, 1998, pp. 541–544.
 25. Алдошина И., Основы психоакустики. Субъективные и объективные методы оценки разборчивости речи [Электронный ресурс], Журнал «Звукорежиссер», 2002, Режим доступа: http://rus.625-net.ru/audioproducer/2002/05/theory_3.htm, Дата доступа: 25.05.2006.
 26. Разработать процедуры сжатия речевой информации, обеспечивающие коммерческое качество восстановленной речи: отчет о НИР (заключительный), БГУИР; рук. темы А.А. Петровский, Минск, 2005, 145 с., № ГР 2002518.
 27. ITU-T Recommendation G.722 «7-kHz audio - coding within 64 kbit/s», 1993.
 28. ITU-T Recommendation G.722.1 «Coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss», September, 1999.
 29. ISO/IEC, JTC1/SC29/WG11 MPEG, "Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s – Part 3: Audio", IS11172-3, 1992. ("MPEG-1").
 30. Wang S., Sekey A., Gersho A., An objective measure for predicting subjective quality of speech coders, IEEE Journal on selected areas in communications, 1992, Vol. 10, No. 5, pp. 819–829.
 31. CORPORA Speech Database: [Electronic resource], Prof. S. Grochowski, supported by the KBN grant 8 T11C 023 8, Poznan, 2003, CD-ROM.
 32. Петровский Ал.А., Построение психоакустической модели в области вейвлет-коэффициентов для перцептуальной обработки звуковых и речевых сигналов, Речевые технологии, № 4, 2008, Москва, с. 61-71.
 33. Petrovsky Al., Petrovsky A., Dynamic algorithm transforms for reconfigurable real-time audio coding processor, The proc. of the Intern. Conference on Parallel Computing in Electrical Engineering, PARELEC'02, Warsaw, Poland, Sep. 22-25, 2002, IEEE Press, Los Alamitos, California, 2002, pp. 422-424.
 34. Petrovsky Al.A., Petrovsky A.A., Reconfigurable processing: the solution to audio coding in real-time, The proc. of the 9th International conference "Mixed design of integrated circuits and systems", MIXDES'2002, 20-22 June 2002, Wroclaw, Poland, pp. 629-634.
 35. Ковалгин Ю.А., Вологдин Э.И., Цифровое кодирование звуковых сигналов, Санкт-Петербург, Корона принт, 2004, 240 с.
 36. Colomes C., Schmidmer C., Thiede T., Treurniet W., Perceptual quality assessment for digital audio: PEAQ – the new ITU standard for objective measurement of the perceived audio quali-

ty, In the proc. of the AES 17th Intern. Conference “High-quality audio coding”, Florence, Italy, 1999, pp. 337 – 351.

37. Петровский Ал.А., Объективная оценка качества восстановленного аудиосигнала перцептуальным ПДВП-кодером на базе периферийной модели уха человека, Сборник трудов 5-й Международной научной конференции “Цифровая обработка сигналов и её применение” (DSPA’2003), т.2, Москва, Россия, 2003, с. 123-126.
38. Петровский Ал.А., Субполосное перцептуальное кодирование звуковых и речевых сигналов, Минск, «Бестпринт», 2008, 218 с.

19 АППАРАТНЫЙ УСКОРИТЕЛЬ НА БАЗЕ ПЛИС С ИНТЕРФЕЙСОМ USB 2.0

Станкевич А.В., Петровский Ал.А., Ключ В.Б., Качинский М.В.

АННОТАЦИЯ

В статье рассматривается аппаратный ускоритель к персональному компьютеру (ПК) на базе ПЛИС XC3S400 семейства Spartan III фирмы Xilinx, предназначенный для реализации трудоемких алгоритмов цифровой обработки информации. Особое внимание уделено организации обмена с ПК по интерфейсу USB, приводятся результаты исследования и практические рекомендации по повышению пропускной способности при обмене данными между аппаратным ускорителем и ПК по USB интерфейсу.

Ключевые слова - аппаратный ускоритель, интерфейс USB, ПЛИС.

СОДЕРЖАНИЕ

19.1 Выбор варианта реализации аппаратного ускорителя	430
19.2 Структура аппаратного ускорителя	431
19.3 Программное обеспечение	434
19.4 Исследование скорости обмена по интерфейсу USB 2.0	435
19.5 Заключение.....	437
19.6 Список использованных источников.....	437

19.1 Выбор варианта реализации аппаратного ускорителя

Одним из наиболее перспективных направлений повышения производительности вычислительных систем является параллельная обработка информации. С этой точки зрения представляет значительный интерес разработка разнообразных аппаратных ускорителей к персональному компьютеру (ПК) для реализации некоторых трудоемких алгоритмов обработки информации. Такими задачами могут быть алгоритмы аудио и видео кодирования и декодирования, шифрования и т. п. Аппаратный ускоритель представляет собой специализированный процессор (СП), реализующий соответствующий алгоритм обработки данных, который подключается к ПК с помощью высокоскоростного интерфейса.

Сформулируем основные требования, которые предъявляются к подобным устройствам и их программному обеспечению:

1) аппаратный ускоритель должен быть специализирован под решение конкретной задачи для достижения максимально возможной производительности;

2) должна иметься возможность изменения алгоритмов обработки данных без переделки конструктивно законченного процессора. Данное требование важно для сокращения сроков разработки новых СП и возможности их перенастройки под другие задачи;

3) аппаратный ускоритель должен иметь высокопроизводительный интерфейс для обмена данными с ПК;

4) должна иметься возможность автоматического обнаружения средствами операционной системы факта подключения к ПК аппаратного ускорителя и его отключения;

5) должны быть реализованы библиотеки пользовательских функций, позволяющие прикладным программам запрашивать требуемые функции специализированных вычислений независимо от того, подключен ли к ПК аппаратный ускоритель или нет. Если ускоритель подключен, то такие функции реализуются аппаратно СП, если нет, то программным путем на ПК.

Рассмотрим возможные варианты реализации аппаратной платформы такого устройства.

Аппаратный ускоритель может быть реализован на базе высокопроизводительного микропроцессора или цифрового сигнального процессора (ЦСП). В этом случае выполнение указанного выше второго требования обеспечивается путем изменения содержимого памяти программ. Другим возможным решением является использование ПЛИС с архитектурой FPGA (field programmable gate array). Данные устройства позволяют изменять алгоритмы обработки данных путем изменения конфигурационной последовательности, загружаемой в ПЛИС после включения питания (загрузкой другого проекта цифровой системы).

Второе решение представляется более предпочтительным, по причине возможности организации параллельной и конвейерной обработки данных, что позволяет при более низкой тактовой частоте синхронизации получить более высокую производительность системы по сравнению с сопоставимыми по стоимости решениями на базе ЦСП и универсальных микропроцессоров.

Имеется достаточно большое количество разнообразных плат для создания прототипов цифровых устройств на базе ПЛИС, которые могут быть использованы в качестве платформы для построения аппаратного ускорителя, однако такие устройства в большинстве случаев обладают аппаратной избыточностью (наличием разнообразных интерфейсов, средств для ввода-вывода аналоговых сигналов и т.д.). Критерий минимизации аппаратных затрат для реализации аппаратного ускорителя важен как с точки зрения уменьшения потребляемой мощности, габаритов и массы, так и с точки зрения снижения стоимости разрабатываемого устройства.

19.2 Структура аппаратного ускорителя

Современные тенденции развития информационных технологий не только определяют появление новых задач и приложений, но, как показывает практика, уже известные задачи иногда получают более эффективное решение на новом технологическом уровне.

Некоторые рекомендации по разработке и программированию ускорителей на базе ПЛИС:

1) уровень реального времени прикладных IP-ядер должен реализовывать параллельно-конвейерную схему обработки потока данных с минимальной временной задержкой между ступенями конвейера;

2) обработка потока данных внутри прикладных IP-ядер должна происходить с интенсивностью, не меньшей интенсивности их поступления;

3) должны быть приняты меры по исключению переполнения буферов входящих/выходящих пакетов данных;

4) реализация должна быть легко масштабируемой и обеспечивать соответствующий прирост производительности вычислительной системы при увеличении аппаратных ресурсов.

В качестве ПЛИС для аппаратного ускорителя выбран кристалл XC3S400 семейства Spartan III фирмы Xilinx [1]. Данное семейство характеризуется невысокой стоимостью и предназначено для построения разнообразных цифровых систем. Кристалл XC3S400 выбран по соображениям максимальной логической емкости (400 000 условных логических вентилей) в корпусе с четырехрядным расположением выводов (не BGA).

Для аппаратного ускорителя весьма важным является выбор интерфейса для обмена данными с ПК, поскольку в большинстве применений именно скорость обмена по интерфейсу сопряжения с ПК определяет общую производительность устройства. Требованиям 3 и 4 предыдущего раздела удовлетворяют интерфейсы USB 2.0 High speed (до 480 Мбит/с) и Ethernet (до 1 Гбита/с). Ethernet обладает большей пропускной способностью, однако USB имеет выходы питания, что позволяет реализовать автономное устройство с питанием от ПК. Кроме того, важным является вопрос разработки программного обеспечения для ПК, которое будет поддерживать обмен данными между аппаратным ускорителем и ПК. С точки зрения уменьшения сроков разработки программного обеспечения был выбран интерфейс USB.

Структура аппаратного ускорителя приведена на рис. 19.1.

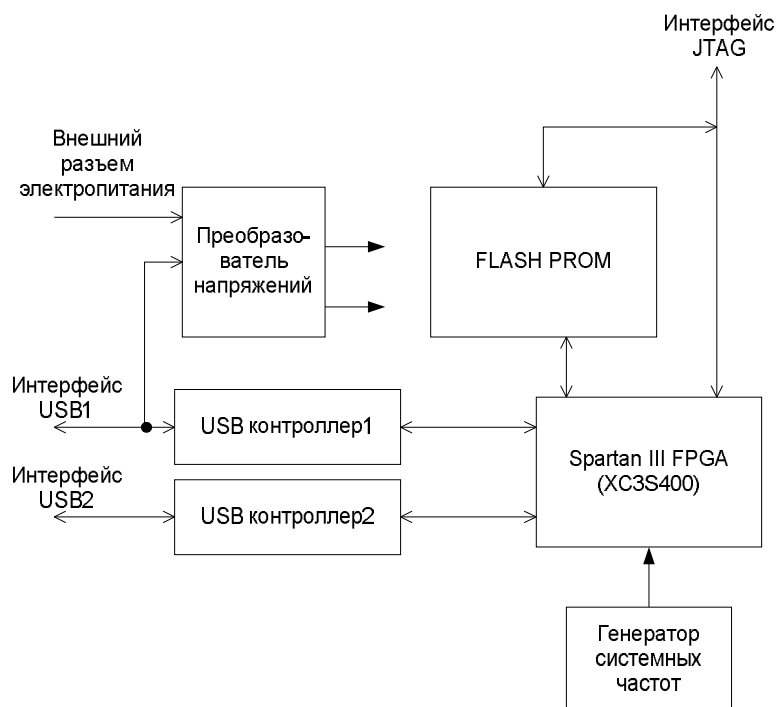


Рис. 19.1 – Структура аппаратного ускорителя.

FLASH PROM необходима для хранения конфигурационной последовательности ПЛИС. Конфигурационная последовательность заносится в PROM через интерфейс JTAG с помощью стандартной утилиты конфигурирования iMPACT пакета ISE Xilinx. Дополнительно в цепочку

сканирования JTAG включен кристалл ПЛИС, что при необходимости позволяет его конфигурировать прямо от ПК.

Преобразователь напряжений предназначен для получения требуемых номиналов напряжений электропитания аппаратного ускорителя. Входным напряжением преобразователя может быть либо напряжение питания интерфейса USB, либо напряжение от внешнего блока питания.

Генератор системных частот предназначен для генерирования частоты синхронизации ПЛИС. Частота синхронизации интерфейса USB формируется через ПЛИС.

В структуре аппаратного ускорителя предусмотрено два контроллера интерфейса USB. Второй контроллер зарезервирован для случая, если через один интерфейс невозможно будет получить требуемую скорость обмена данными с аппаратным ускорителем. В этом случае один интерфейс USB будет использоваться для приема, а второй – для передачи данных.

В качестве контроллеров USB выбран CY7C68001 фирмы Cypress [2]. Основными аргументами для выбора этого контроллера были его популярность и то, что Cypress предоставляет пользователю законченный комплекс USBDevStudio программных утилит и драйвер для операционной системы Microsoft Windows, значительно снижающие сроки разработки.

Для уменьшения числа корпусов на плате функции инициализации и записи дескриптора в контроллер CY7C68001 были возложены на ПЛИС.

Структура блока сопряжения СП с контроллерами USB, реализуемого как часть проекта ПЛИС, приведена на рис. 19.2. Данная структура приведена для случая, когда интерфейс USB1 используется для приема данных в аппаратный ускоритель, а интерфейс USB2 используется для передачи данных от аппаратного ускорителя.

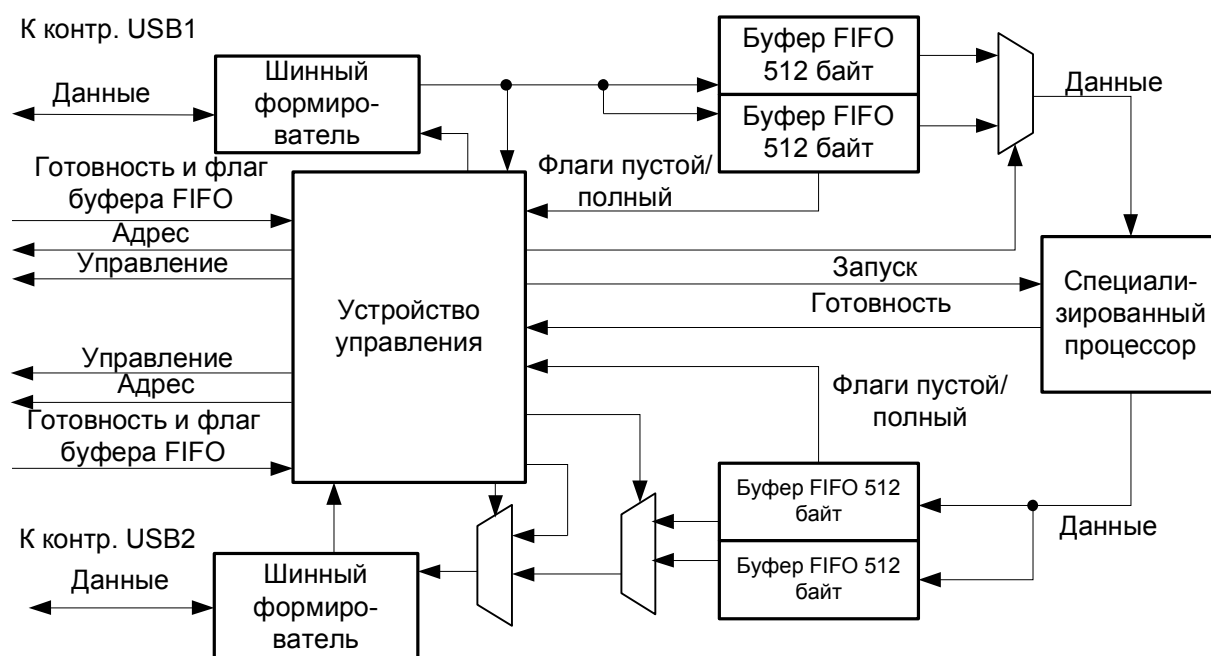


Рис. 19.2 – Структура блока сопряжения СП с контроллерами USB.

Входные данные для СП поступают от ПК на контроллер USB1. Далее через двунаправленные формирователи данные поступают на буферы FIFO. Два приемных буфера FIFO предназначены для совмещения во времени процессов занесения от интерфейса USB очередного бло-

ка данных из 512 байт и формирования данных для СП. Далее данные поступают на СП, который реализует требуемый алгоритм обработки данных.

Для запуска процесса обработки используется сигнал «Запуск» от устройства управления. По завершении преобразования данных СП выдает сигнал «Готовность». Выходные данные СП переписываются в один из свободных выходных буферов FIFO. Наличие двух буферов объясняется необходимостью совмещения во времени процессов записи в буфер выходных данных от СП и передачи данных по интерфейсу USB. Далее данные через мультиплексоры и шинные формирователи передаются контроллеру USB, выполняющему функции передатчика.

Устройство управления обеспечивает формирование сигналов, управляющих работой остальных блоков, а также начальное конфигурирование контроллеров USB.

Внешний вид аппаратного ускорителя приведен на рис. 19.3.

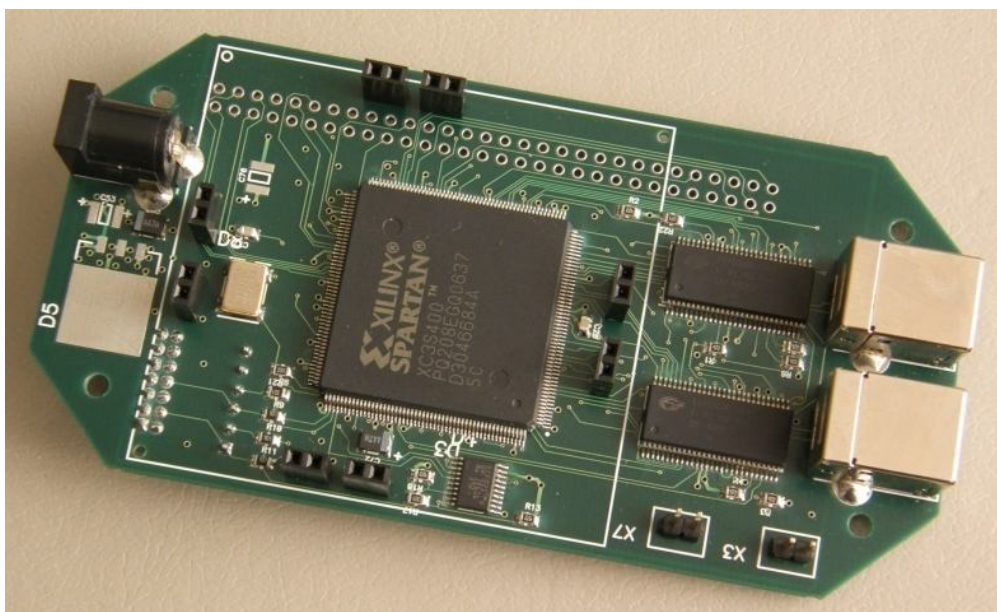


Рис. 19.3 – Внешний вид аппаратного ускорителя.

Аппаратные затраты ресурсов кристалла XC3S400 для реализации блока сопряжения СП с контроллерами USB приведены в табл. 19.1.

Табл. 19.1 – Аппаратные затраты ресурсов кристалла XC3S400 для реализации блока сопряжения специализированного процессора с контроллерами USB.

Ресурс	Использовано	Процент от общего числа
Slice Flip Flops	369	5,1%
LUT	627	8,7%
Slice	512	14,3%
BRAM	2	12,5%

19.3 Программное обеспечение

Разработанный аппаратный ускоритель содержит два оконечных устройства для подключения к шине USB. Первое оконечное устройство служит для приема информации от ПК, а

второе – для передачи в ПК. Каждое из оконечных устройств, подключенных к ПК по шине USB, требует наличия двух оконечных точек с различными типами передачи: CONTROL – для идентификации оконечного USB устройства USB хост контроллером и установления драйвера в операционной системе и BULK – для приема или передачи информации. Выбор типа обмена данными BULK [3] по интерфейсу USB обусловлен такой особенностью, как гарантированная доставка данных без ошибок в отличие от других типов обмена. Для обеспечения высокой скорости передачи данных каждое из оконечных USB устройств конфигурирует USB хост контроллер на высокоскоростной режим передачи.

Типовая структура программного модуля для взаимодействия с аппаратным ускорителем на ПК включает следующие уровни:

- 1) приложение пользователя;
- 2) библиотека функций работы с модулем аппаратного ускорителя USBlib, которая включает в себя библиотеку работы с USB интерфейсом CyAPI;
- 3) драйвер USB устройства CyUSB.sys;
- 4) драйвер USB хост контроллера (штатный компонент операционной системы Microsoft Windows).

19.4 Исследование скорости обмена по интерфейсу USB 2.0

Для исследования скорости обмена по интерфейсу USB рассмотрим различные способы подключения аппаратного ускорителя к ПК (рис. 19.4).

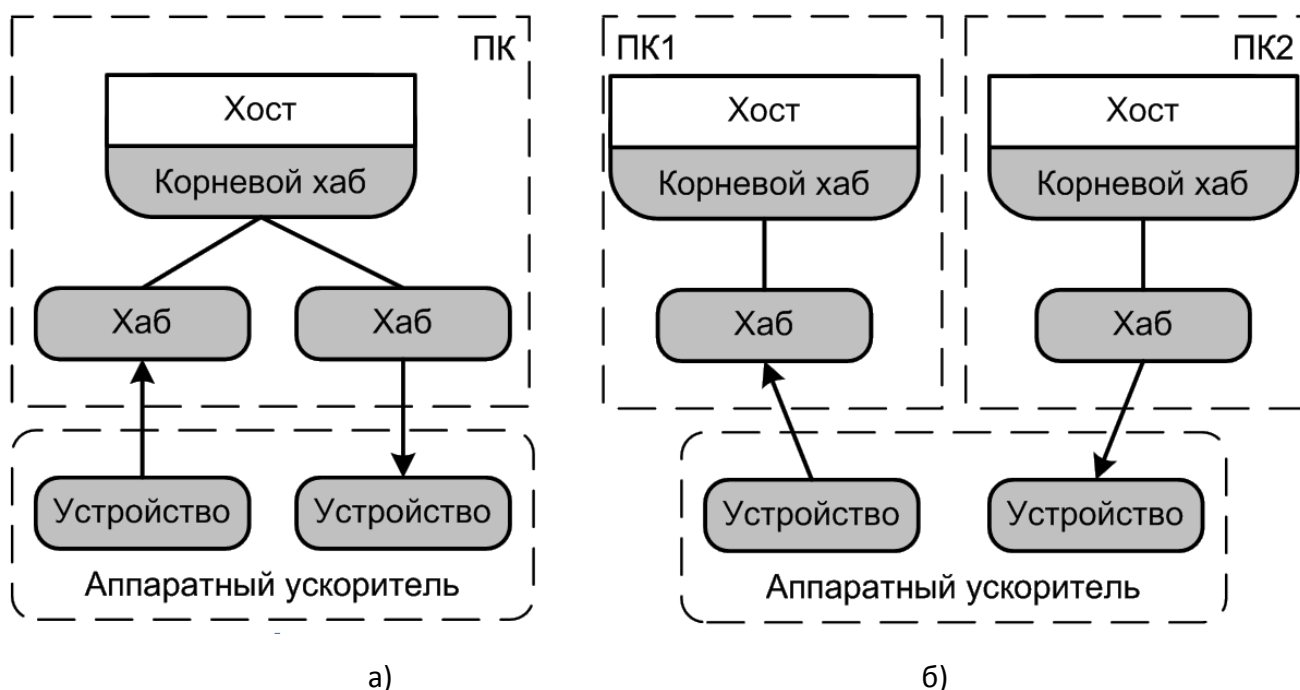


Рис. 19.4 – Схемы подключения аппаратного ускорителя к шинам USB: (а) - одного ПК; (б) - разных ПК

Первая - типичная схема подключения нескольких оконечных устройств к одной шине USB одного ПК (рис. 19.4а). В этом случае два оконечных USB устройства аппаратного ускорителя

ля подключаются к шине USB через хабы, как два независимых оконечных устройства. В данном случае все ресурсы USB хост контроллера будут отданы одному из устройств, если оно активно или разделены между устройствами, если они оба активны в текущий момент времени.

Вторая - схема подключения нескольких оконечных устройств к шине USB разных ПК (рис. 19.4б). В такой схеме два оконечных USB устройства аппаратного ускорителя подключаются к шине USB через хаб каждого отдельного ПК, как независимые оконечные устройства. В данном случае все ресурсы USB хост контроллера каждого из ПК будут отданы соответствующему устройству, если оно активно в текущий момент времени.

Тесты для проверки пропускной способности заключались в передаче от ПК большого файла (размером более 40 Мбайт), сохранении принимаемых данных во втором файле и сравнении содержимого первого и второго файла.

При передаче замеряется время, затраченное на выполнение передачи данных. Пропускная способность рассчитывается как размер исходного файла, деленный на время передачи данных. При этом такой тест должен быть выполнен не менее 3 раз. За оценку пропускной способности должно быть взято среднее значение. Усреднение необходимо для учета многозадачности работающей операционной системы. Также реализовывались тесты для односторонней передачи и одностороннего приема данных. Исследования проводились на разных аппаратных конфигурациях, в частности:

1) материнская плата Gigabyte 8IPE100-G/L, процессор Intel Pentium 4 2.4 ГГц, память 2x512 Мб, HDD 80 Гб ST380011A, операционная система Microsoft Windows XP SP3;

2) компьютер Compaq HP 2510p на базе процессора Intel Core 2 Duo, чипсет U7600 1.2 ГГц, память 2 Гб, HDD 100 Гб Toshiba MK1011GAN, операционная система Microsoft Windows XP SP3.

Результаты тестирования сведены в табл. 19.2.

Табл. 19.2 – Результаты тестирования работы аппаратного ускорителя для определения пропускной способности по интерфейсу USB.

№	Эксперимент	Скорость передачи, Мбит/с
1	Передача данных от USB порта ПК к оконечному USB устройству аппаратного ускорителя	235-260*
2	Прием данных от оконечного USB устройства аппаратного ускорителя в USB порт ПК	250-300*
3	Передача и прием для одного ПК в соответствии со схемой рис. 4, а	90-170*
4	Передача от ПК1 и прием в ПК2 в соответствии со схемой рис. 4, б	220-250*

*Разброс скорости передачи, определяется различными аппаратно-программными конфигурациями ПК.

Из приведенных данных следует, что ресурсов одного ПК недостаточно для обеспечения двухстороннего обмена данными с аппаратным ускорителем на скоростях более 170 Мбит/с.

Эксперимент с подключением разработанного устройства к двум USB портам разных ПК показывает возможность аппаратного ускорителя обеспечивать пропускную способность более 200 Мбит/с.

В качестве практических рекомендаций по повышению пропускной способности при обмене данными между аппаратным ускорителем и ПК по USB интерфейсу можно предложить следующие:

1) для обмена лучше использовать порты хост-контроллера USB, расположенного на материнской плате ПК;

2) перед организацией обмена следует отключить от ПК все работающие USB устройства, поскольку они все осуществляют обмен через один хост-контроллер и один системный драйвер.

19.5 Заключение

Аппаратные ускорители, построенные на базе современных кристаллов ПЛИС, позволяют создавать сложные цифровые системы с поддержкой высокоскоростных интерфейсов. Высокоуровневые средства разработки аппаратно-программных решений, а также большое число проектов с открытым исходным кодом позволяют значительно уменьшить сроки создания конечных продуктов на ПК. Имея в своей системе ускоритель на базе ПЛИС, пользователь получает не просто ускоритель вычислительных функций, но и эффективно расширяемую и настраиваемую аппаратную платформу.

19.6 Список использованных источников

1. Кузелин М.О., Кнышев Д.А., Зотов В.Ю., Современные семейства ПЛИС фирмы Xilinx, М., Горячая линия – Телеком, 2004.
2. Долгушин С., Высокоскоростные контроллеры USB производства компании Cypress, Компоненты и технологии, 2006, № 6, с. 25.
3. Чекунов Д., Программисту USB-устройств. Знакомство с USB, Современная электроника, 2004, Октябрь, с. 68–71.

20 ПДВП-АНАЛИЗАТОР ЗВУКОВЫХ СИГНАЛОВ НА АППАРАТНОМ УСКОРИТЕЛЕ НА БАЗЕ ПЛИС

Петровский Ал.А., Станкевич А.В.

АННОТАЦИЯ

Показана возможность реализации ПДВП-анализатора звуковых сигналов на базе ПЛИС. В основу вычислительного ядра положена МАС-структура. Процессор представляет собой параллельно-поточную структуру для дерева, аппроксимирующего критический частотно-временной план. Для реализации анализатора на базе ПЛИС XC3S200 фирмы Xilinx использовано около 40% кристалла по логическим блокам, 50% блочной памяти, 66% матричных умножителей. Оставшего ресурса кристалла достаточно для реализации синтезатора. Моделирование работы анализатора в режиме *post place and route* показало, что проект будет работоспособен до тактовой частоты 70 МГц.

Ключевые слова - ПЛИС, МАС-структура, ПДВП.

СОДЕРЖАНИЕ

20.1 Введение.....	438
20.2 Аппроксимация критического частотно- временного плана	440
20.3 ПДВП-процессор реального времени	441
20.3.1 Функциональная схема вычисления алгоритма ПДВП.....	441
20.3.2 Реализация базовой операции ПДВП на МАС структуре.....	443
20.4 Параллельно-поточная архитектура процессора ПДВП.....	445
20.5 Реализация ПДВП-анализатора на ПЛИС.....	448
20.5.1 Блок-схема ПДВП_анализатора на ПЛИС XC3S200 фирмы Xilinx.....	448
20.5.2 Структура блока сопряжения с DSP....	449
20.5.3 Вычислительное ядро	450
20.5.4 Структура памяти ступени анализатора.....	452
20.6 Заключение.....	453
20.7 Список использованных источников.....	454

20.1 Введение

Структуры большинства кодеров звука на основе психоакустики сходны и могут быть представлены обобщенной схемой (см. рис. 20.1) [1].

Входной звуковой сигнал $x(n)$ разбивается на квазистационарные фреймы длительностью от 2 до 50 миллисекунд в зависимости от алгоритмов обработки и методов кодирования.

Блок частотно-временного анализа аппроксимирует временные и спектральные параметры звука для каждого анализируемого фрейма с учетом шкалы критических частотных полос. В блоке психоакустического анализа оценивается энергия маскирующего сигнала (уровни маскирующих порогов) на базе психоакустической модели. При этом определяются максимальные искажения, возникающие в каждой точке частотно-временной плоскости в процессе квантования и кодирования частотно-временных оценок без введения искусственного артефакта слышимости при восстановлении сигнала. Следовательно, психоакустический анализатор вычисляет частотно-временной параметр невосприятия акустической информации слушателем, который затем передается в блок квантования и кодирования. В процессе психоакустического кодирования необходимо: во-первых, установить вид маскирующего сигнала, во-вторых, вычислить соответствующие пороги, затем полученную информацию использовать для того, чтобы расположить спектр шума кодирования ниже порога маскирования, известный как порог едва различимых искажений *JND* (just noticeable distortion) [2].

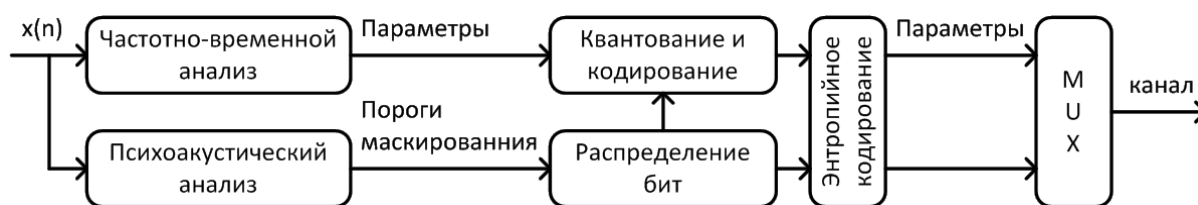


Рис. 20.1 – Общая структура перцептуального кодера звука

Выбор метода частотно-временного анализа неразрывно взаимосвязан с точностью вычисления психоакустической модели и общей концепцией построения системы кодирования сигналов звука. При этом выбор методологии частотно-временного анализа всегда обуславливает решение фундаментального вопроса: обеспечения требуемого разрешения как в частотной так и временной областях. Развитие этих работ в настоящее время идет по пути построения перцептуальных кодеров звука на основе подходящего частотно-временного преобразования, в частности, банков цифровых фильтров и наиболее гибкого, в смысле управления частотно-временным разрешением, быстрого вейвлет преобразования (пакет дискретного вейвлет преобразования (ПДВП)) с динамической трансформацией дерева алгоритма.

ПДВП есть обобщение диадического вейвлет преобразования, которое позволяет получить множество структур путем его соответствующих декомпозиций [3]. Структура ПДВП больше согласуется с сигналом, чем вейвлет преобразование [4], и характеризуется следующими свойствами: малая вычислительная сложность процедуры декомпозиции звукового сигнала в выбранном базисе (процедура анализа); малая вычислительная сложность процедуры реконструкции сигнала (синтеза) в выбранном базисе; конвейерность вычислительного процесса процедур анализа и синтеза, что способствует организации поточных и параллельно-поточных структур процессоров реального времени; гибкое изменение частотно-временного разрешения, обеспечивающее локализацию нестационарностей в сигнале; единственность преобразования, т.е. в ограниченном числе структур ПДВП имеется одна структура, идентифицирующая соответствующие компоненты сигнала [5].

20.2 Аппроксимация критического частотно-временного плана

Аппроксимация критической шкалы частот на основе ПДВП осуществляется таким образом, чтобы расстояние между центральными частотами $z(f)$ полос пропускания было размером в один барк. Соотношение между шкалой частот в герцах и барках приведено в [6].

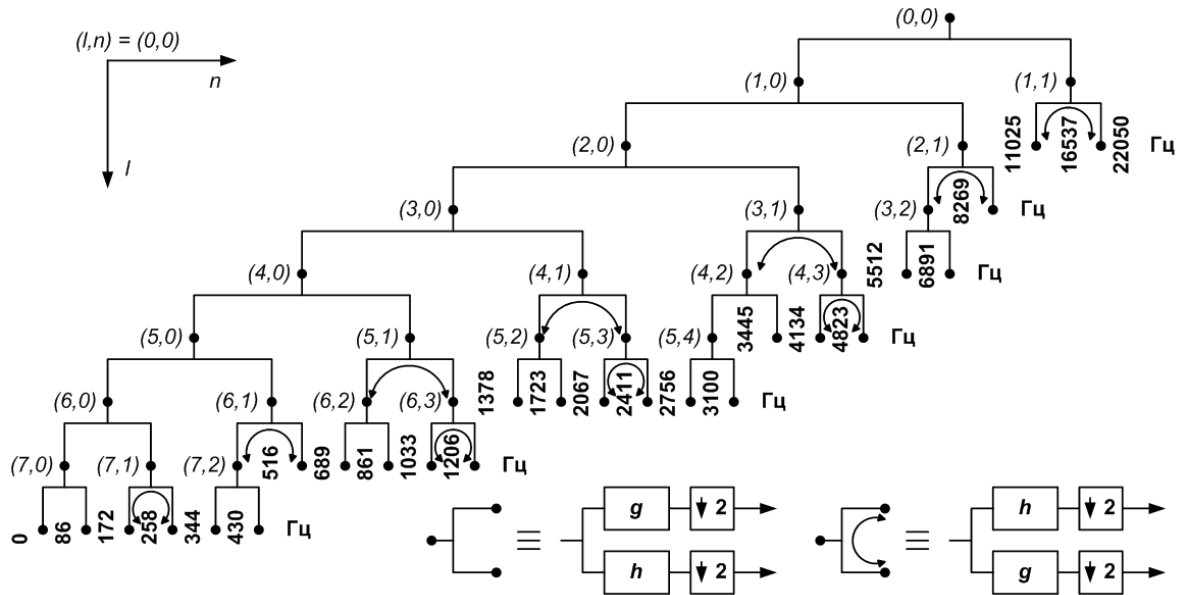


Рис. 20.2 – Дерево ПДВП $(l, n) \in E_{CB}$

На рис. 20.2 показано дерево ПДВП (Critical Band Wavelet Packet Decomposition ($CB - WPD$)), полученное эмпирически, которое осуществляет разделение частотного интервала звукового сигнала на полосы, согласно критической шкале частот [7]:

$$CB - WPD: (l, n) \in E_{CB}, l = \overline{0,8},$$

где E_{CB} – обозначает множество узлов дерева ПДВП соответствующего $CB - WPD$.

На рис. 20.3 и рис. 20.4 показаны аппроксимации центральной частоты и ширины каждой частотной полосы критической шкалы частот соответственно деревом ПДВП, структура которого приведена на рис. 20.2.

Дерево $CB - WPD$ делит частотный диапазон $[0 - 22,05 \text{ кГц}]$ на 25 неравномерных полос $CBW(f)$ согласно [6], т.е. на 25 барков. Корневой узел $(l, n) = (0,0)$ данного дерева соответствует всему частотному диапазону сигнала. Каждый внутренний узел дерева $(l, n) \in E$, названный узлом предка, делится на два потомка: 1-й потомок и 2-й потомок, ассоциируемые соответственно с высокочастотной и низкочастотной фильтрацией, выходные сигналы (вейвлет коэффициенты) которых децимируются в соотношении 2:1:

$$X_{l,n,k}(t) = \langle x(t), \varphi_{l,n,k}(t) \rangle, (l, n) \in E_{CB}, k \in \mathbb{Z}.$$

где l - номер масштабного уровня преобразования ($0 \leq l \leq 8$), n - номер узла масштабного уровня преобразования, k – вейвлет коэффициентов в полосе (узле (l, n) дерева E). Банк вейвлет фильтров ($CB - WPD: (l, n) \in E_{CB}$), согласованный с критической шкалой частот восприятия акустической информации человеком, является предельной структурой для метода перцептуального кодирования сигнала звука [7].

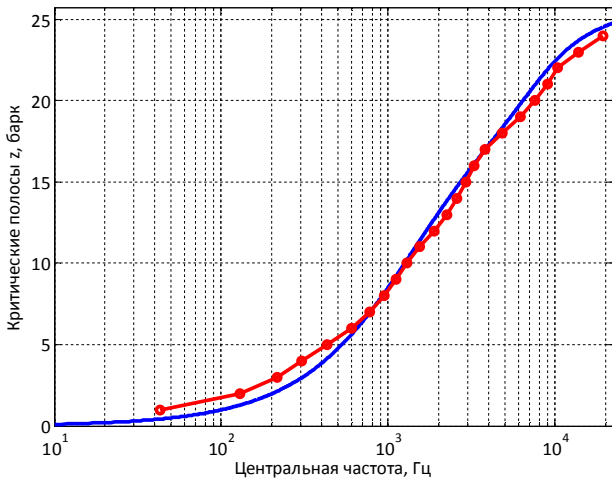


Рис. 20.3 – Аппроксимация центральных частот $CB - WPD: (l, n) \in E_{CB}, l = \overline{0,8}$

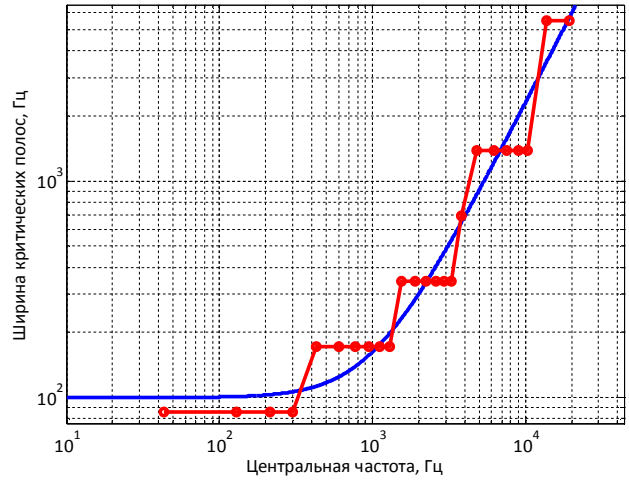


Рис. 20.4 – Аппроксимация ширины критических частотных полос $CB - WPD: (l, n) \in E_{CB}, l = \overline{0,8}$

20.3 ПДВП-процессор реального времени

20.3.1 Функциональная схема вычисления алгоритма ПДВП

Реализация ПДВП-процессора для полного дерева ПДВП предложена в [8], а для перцептуально адаптированного дерева ПДВП показана в [9]. Создание ПДВП-процессора способного адаптировать ПДВП под текущий фрейм сигнала требует разработки нового подхода. Рассмотрим пример функциональной схемы вычисления адаптивного ПДВП для трех масштабных уровней показан на рис. 20.5. Здесь приняты следующие обозначения: ПМ – процессорный модуль базовой операции ПДВП, а МБП – модуль буферизации и переключения. МБП (рис. 20.6) состоит из ОЗУ, переключателя SW , управляемого параметрами α_l, β_n вектора реконфигурации процессора $r_{l,n}$. МБП необходим для хранения промежуточных результатов и их вывода для обработки в DSP процессоре.

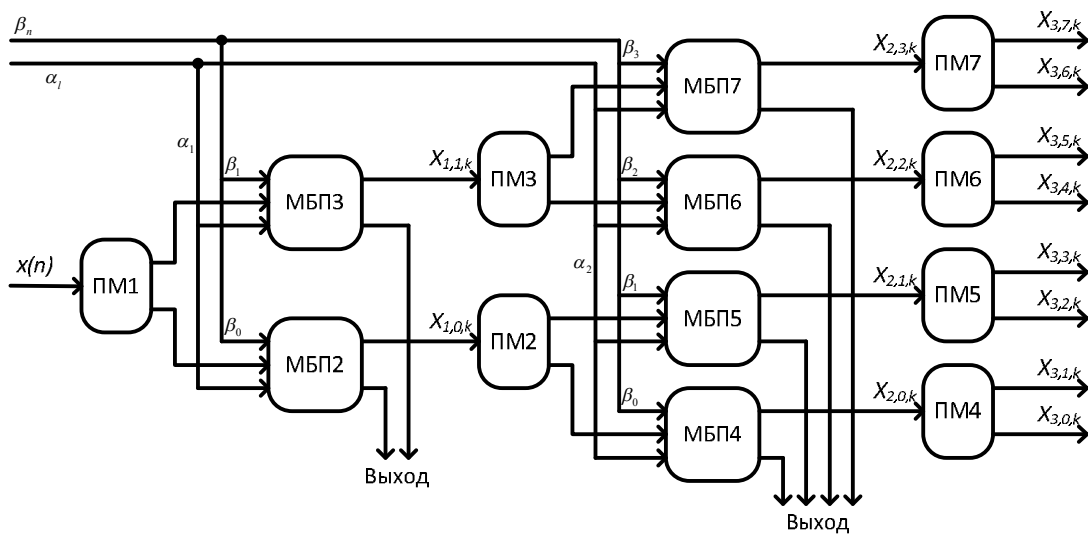


Рис. 20.5 – Функциональная схема вычисления адаптивного ПДВП

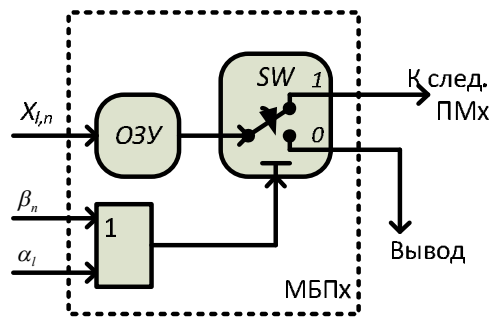


Рис. 20.6 – Структура модуля буферизации и переключения

Положение переключателя SW в зависимости от параметров α_l, β_n приведено в табл.

20.1.

Табл. 20.1 – Таблица переключателя SW

Биты параметров вектора $r_{l,n}$		Состояние переключателя SW	Направление потока данных
α_l	β_n		
0	0	0	Вывод данных
1	0	0	Вывод данных
0	1	0	Вывод данных
1	1	1	Переход к след. ПМх

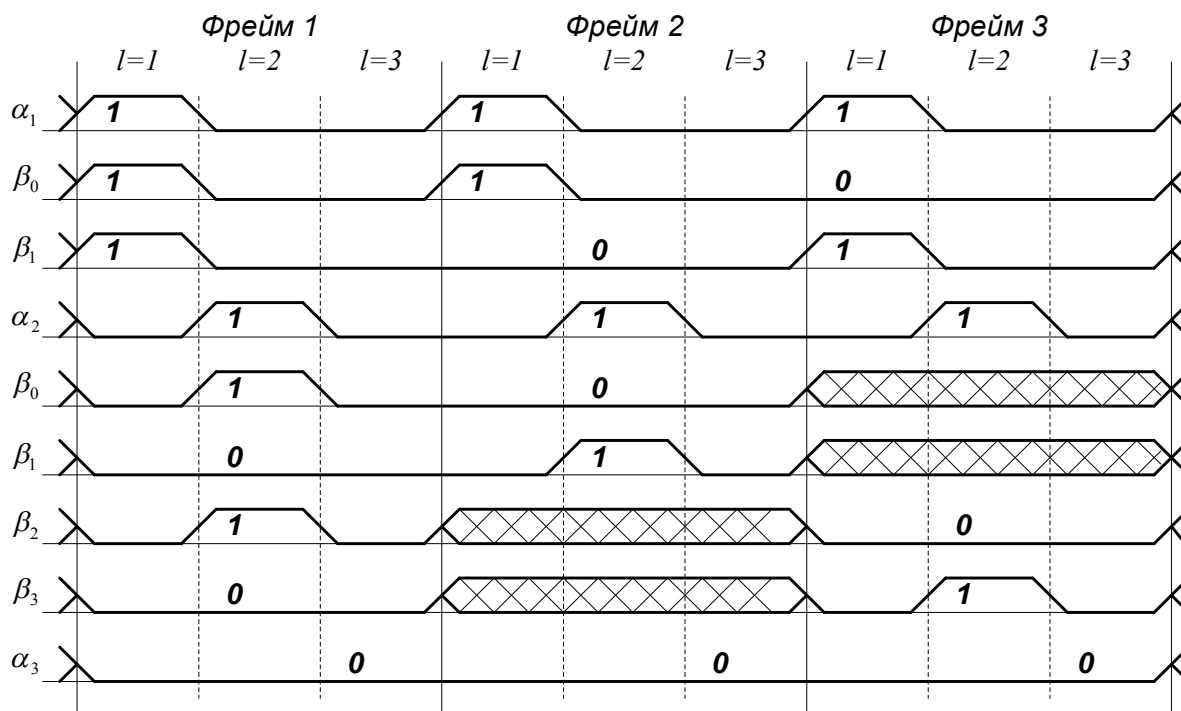


Рис. 20.7 – Временная диаграмма изменения управляющих сигналов для трех последовательных фреймов звукового сигнала

Диаграмма изменения во времени параметров α_l, β_n вектора трансформации процессора $r_{l,n}$ в соответствии со структурой дерева $(l, n) \in E_i$ для трех последовательных фреймов звукового сигнала приведена на рис. 20.7. Структуры деревьев ПДВП, отображенные пунктир-

ной линией на рис. 20.8 для соответствующих фреймов, определяют варианты их дальнейшего роста в соответствии с полученными значениями перцептуальной энтропии $PE_{l,n}$ в каждом узле дерева, определяющее информативную составляющую полученной декомпозиции дерева ПДВП, показывает неэффективность дальнейшего роста структуры дерева.

Таким образом, структуры деревьев ПДВП E_i , описываемые узлами (l, n) , а также соответствующие вектора реконфигурации процессора следующие:

- для 1-го фрейма $E_1 = \{(1,0); (1,1)\}, \{(2,0); (2,1); (2,2); (2,3)\}$ и соответствующий вектор $r_1 = [(1,1,1), (1,1,0,1,0), (0)]$;
- для 2-го фрейма $E_2 = \{(1,0); (1,1)\}, \{(2,0); (2,1)\}$,
 $r_2 = [(1,1,0), (1,0,1, x, x), (0)]$;
- для 3-го фрейма $E_3 = \{(1,0); (1,1)\}, \{(2,2); (2,3)\}$,
 $r_3 = [(1,0,1), (x, x, 0, 1, 0), (0)]$.

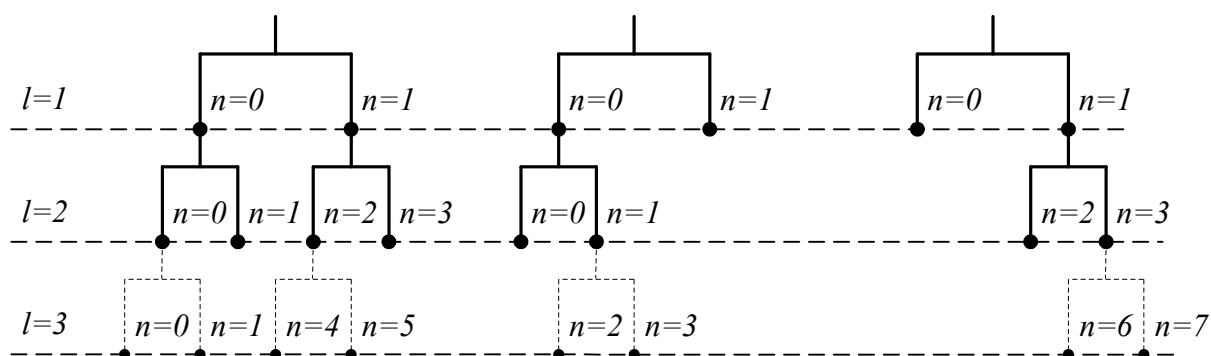


Рис. 20.8 – Пример структуры деревьев ПДВП

20.3.2 Реализация базовой операции ПДВП на МАС структуре

Реализация базовой операции ПДВП на МАС структуре (рис. 20.9): последовательно включены умножитель и накапливающий сумматор (multiplier-accumulator) [9], т.е. прямая реализация выражения:

$$y(k) = \sum_{n=0}^N h(n)x(k - n), \quad (20.1)$$

где $h(n)$ - коэффициенты биортогонального вэйвлет-фильтра; $x(k)$ - входная последовательность фильтра; $y(k)$ - выходная последовательность после фильтрации.

Здесь входная последовательность $x(n)$ буферизируется в блоке «буфер входных отсчетов». Отсчеты буфера $x(n)$ поступают на два высокоскоростных последовательных умножителя, где происходит их умножение на коэффициенты вэйвлет-фильтров (нижних и верхних частот соответственно вэйвлет- $h_1(k)$ и масштабной $h_0(k)$ функцией). Результаты произведений $x(k)h_0(k)$ и $x(k)h_1(k)$ суммируются с предыдущим значением и накапливаются в аккумуляторе. Далее продукт фильтрации поступает в регистр-дециматор, который осуществляет децимацию на 2, т.е. частота дискретизации f_s уменьшается в два раза $f_s/2$. Данные регистры также играют роль выходного буфера. Необходимое количество КЛБ для реализации базовой операции ПДВП на МАС структуре приведено в табл. 20.2.

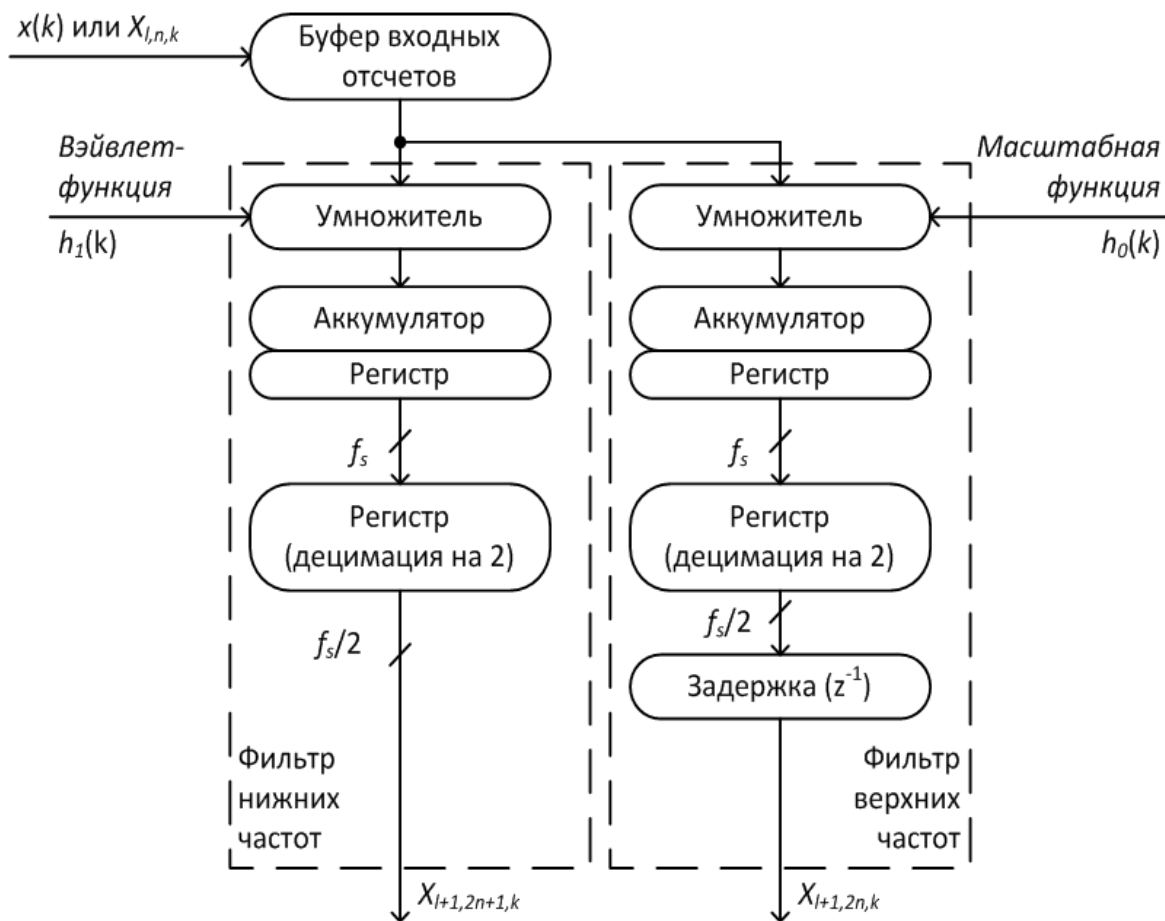


Рис. 20.9 – Реализация базовой операции ПДВП на MAC структуре.

На данной MAC структуре можно эффективно реализовывать вэйвлет-фильтры высокого порядка, например, вэйвлет-фильтры семейства Добеши 20-го порядка (40 коэффициентов). В данном случае для построения базовой операции ПДВП потребуется не более 1.5 КЛБ на порядок фильтра. Скорость потока поступления входных данных при этом сохраняется приблизительно от 100 до 1000 кГц, что вполне достаточно для звуковых сигналов с частотами дискретизации 44.1 кГц и 48 кГц.

Табл. 20.2 – Необходимое количество КЛБ при реализации базовой операции ПДВП на MAC структуре

Порядок фильтра	Длина слов, бит									
	6	8	10	12	14	16	18	20	22	24
8	20	24	30	34	40	44				
16	38	48	60	68	78	88	98			
32	76	96	116	136	156	176	196	216		
48	114	144	174	204	234	264	294	324	354	
56		168	204	238	274	308	348	378	414	448
64			232	272	312	352	392	432	472	512
96				408	468	528	588	648	708	768
128					624	704	784	864	944	1024

20.4 Параллельно-поточная архитектура процессора ПДВП

Вычисление на каждом масштабном уровне структуры дерева ПДВП выполняется только над данными одного фрейма, а процесс конвейерной обработки может быть организован путем выполнения процесса обработки сразу нескольких фреймов входного сигнала в одно и тоже время, но на разных масштабных уровнях l структуры дерева ПДВП. Временная диаграмма параметров вектора трансформации $r_{l,n}$ для конвейерного процесса обработки нескольких последовательных фреймов входного сигнала на основе алгоритма динамической трансформации структуры дерева ПДВП показана на рис. 20.10.

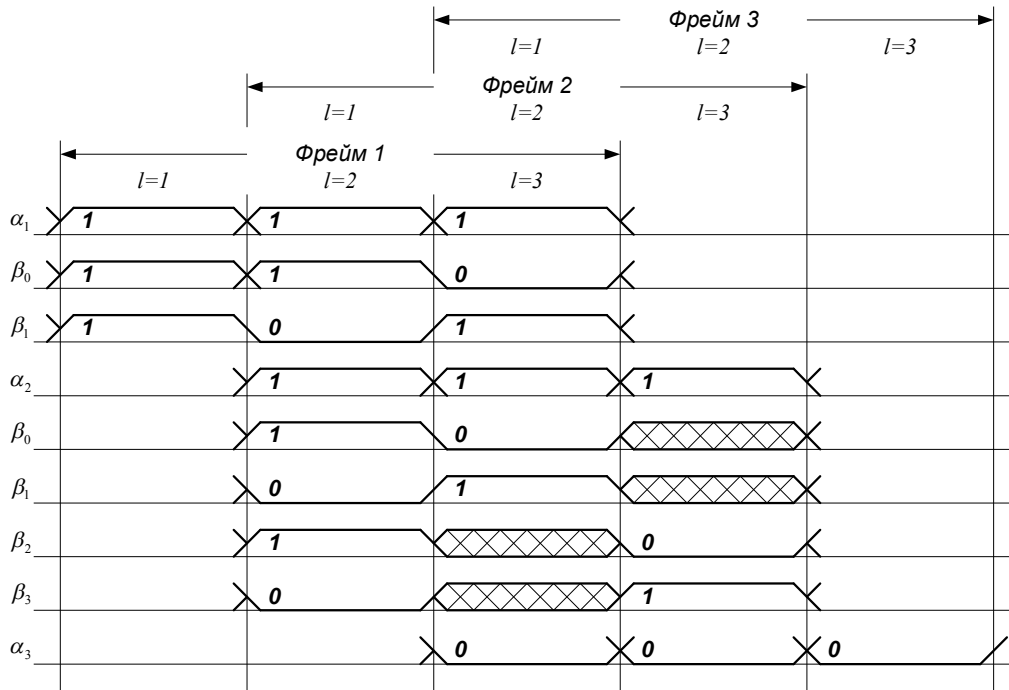


Рис. 20.10 – Временная диаграмма при совмещенной обработке нескольких последовательных фреймов

Примеры структур деревьев ПДВП для каждого фрейма звукового сигнала приведены на рис. 20.11, откуда видно, что формирование структуры дерева ПДВП для каждого фрейма осуществляется на один масштабный уровень позже, чем для предыдущих фреймов входного звукового сигнала.

В соответствии со структурой дерева ПДВП количество ветвей с каждым новым уровнем возрастает вдвое, а поток обрабатываемых данных в ветви уменьшается на такую же величину. Следовательно, обработку входных данных на уровне l можно выполнять на базе одного ПМ базовой операции. При этом, представляется возможность минимизации числа КЛБ путём модификации МБП, делая последний более универсальным.

Как уже отмечалось выше, структура ПДВП-процессора закладывается на предельный случай – для дерева $CB - WPD: (l, n) \in E_{CB}, l = \overline{0,8}$. При этом, объём модулей памяти, их число, количество переключателей и их вид изменяется от одного масштабного уровня структуры дерева ПДВП к другому, в соответствии с табл. 20.3.

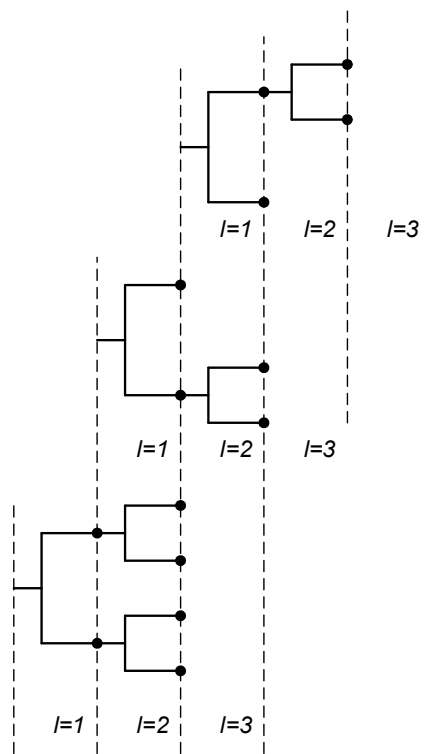


Рис. 20.11 – Пример структур деревьев ПДВП для рис. 20.10

Табл. 20.3 – Перечень элементов МБП в зависимости от масштабного уровня дерева ПДВП

Масштабный уровень	Число элементов ОЗУ и их объём	Число переключателей и их вид
1 (1 ступень)	2 (512×16 бит)	2 (1×2) + 1 (2×1)
2 (2 ступень)	4 (256×16 бит)	4 (1×2) + 1 (3×1)
3 (3 ступень)	4 (128×16 бит)	4 (1×2) + 1 (3×1)
4 (4 ступень)	6 (64×16 бит)	6 (1×2) + 1 (4×1)
5 (5 ступень)	8 (32×16 бит)	8 (1×2) + 1 (5×1)
6 (6 ступень)	10 (16×16 бит)	10 (1×2) + 1 (4×1)
7 (7 ступень)	8 (8×16 бит)	8 (1×2) + 1 (3×1)
8 (8 ступень)	-	-

На рис. 20.12 приведена структура модифицированного МБП в соответствии с табл. 20.3 для шестой ступени. МБП состоит из 5-ти модулей буферизации и переключения, активируемых сигналами управления *ModSel* и *Level* – номер ступени процессора, а также выходного коммутатора 4×1, управляемого вектором трансформации $r_{l,n} = \{\alpha_l, \beta_n\}$. Общая синхронизация осуществляется по сигналу *Sync*.

Архитектура параллельно-поточного процессора ПДВП показана на рис. 20.13, была предложена в [9]. Для каждого масштабного уровня дерева ПДВП используется один ПМ и один МБП, за исключением последнего – восьмого, т.к. архитектура ориентирована на структуру дерева *CB – WPD*. Это объясняется тем, что с ростом структуры дерева ПДВП количество ветвей с каждым новым масштабным уровнем удваивается, скорость поступления данных в этих ветвях соответственно сокращается в два раза.

Структура МБП (рис. 20.12) позволяет, начиная со второго масштабного уровня дерева ПДВП, использовать один и тот же ПМ на ступени параллельно-поточного процессора для обработки поочередно поступающих потоков данных, осуществлять временное хранение, передачу данных для следующей ступени процессора и обеспечение записи результатов обработки в ЦПОС – систему мониторинга.

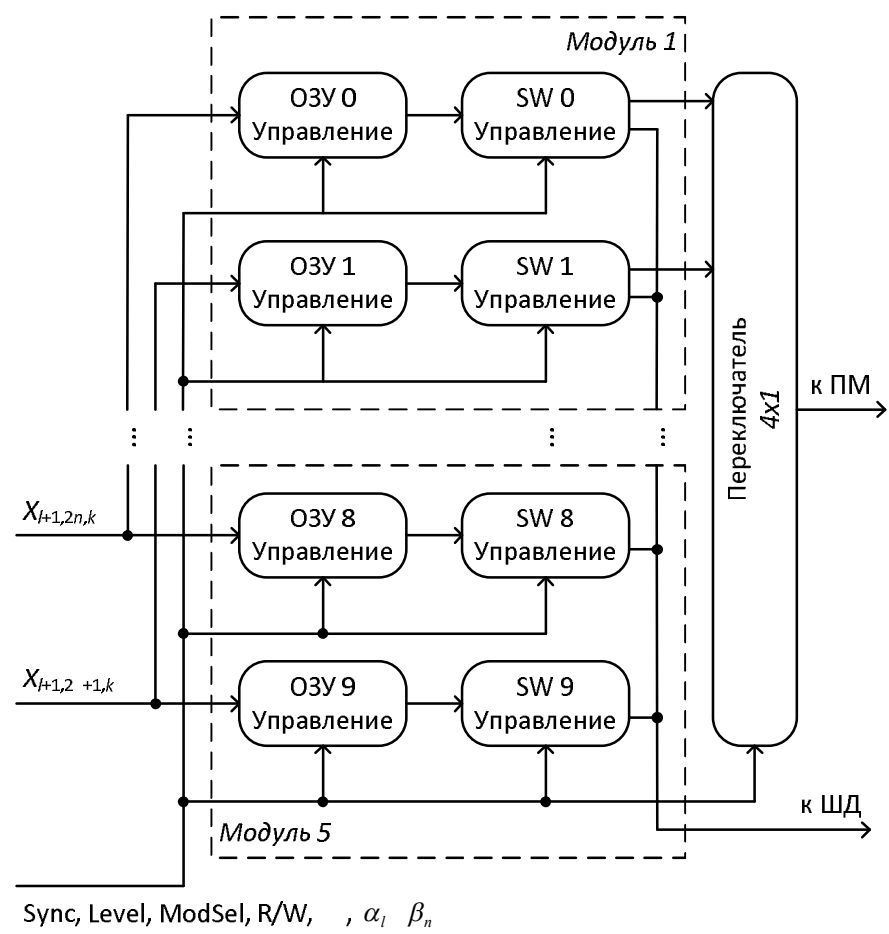


Рис. 20.12 – Модифицированный МБП

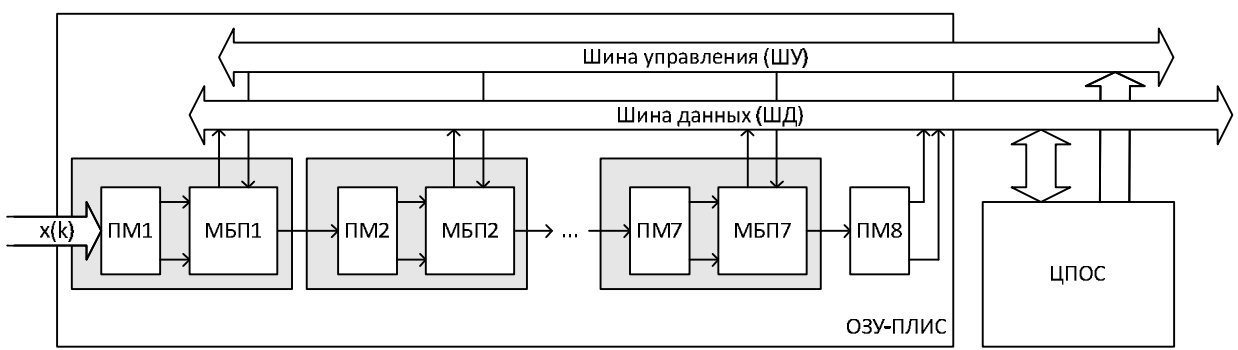


Рис. 20.13 – Архитектура параллельно-поточного ПДВП-процессора

В заключение следует отметить, что, несмотря на ориентацию архитектуры процессора на дерево E_{CB} ПДВП, согласно алгоритму динамической трансформации дерева ПДВП под заданный вычислительный ресурс, на данной реконфигурируемой параллельно-поточной архитектуре ПДВП-процессора возможна структурная минимизация потребляемой энергии процессором за счёт включения/выключения ступеней процессора.

20.5 Реализация ПДВП-анализатора на ПЛИС

20.5.1 Блок-схема ПДВП_анализатора на ПЛИС XC3S200 фирмы Xilinx

Блок анализа реализован на ПЛИС XC3S200 фирмы Xilinx. Разработка проекта велась с помощью интегрированной среды проектирования ISE 7.1 и симулятора ModelSimXE III 6.0.

Укрупненная структура анализатора представлена на рис. 20.14.

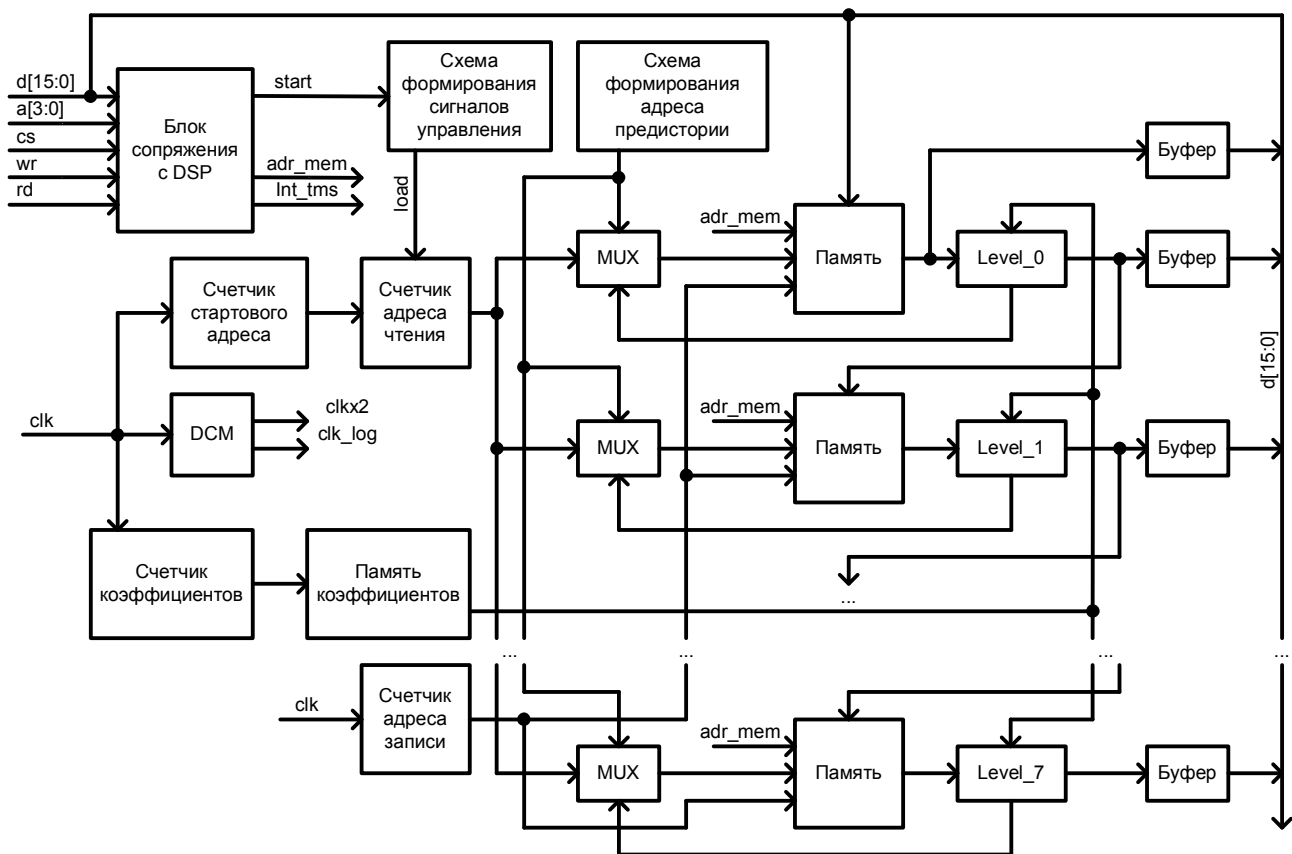


Рис. 20.14 – Структура ПДВП-анализатора на ПЛИС

Блок анализа состоит из 8 одинаковых ступеней, в каждую из которых входит блок памяти объемом 512x16 (512 16-разрядных слов), вычислительное ядро ступени (level_0, ... level_7) и буфер с тремя состояниями для подключения к двунаправленной шине данных DSP (d[15:0]).

Блок памяти каждой ступени представляет собой память, состоящую из двух или более буферов, с тремя независимыми адресными портами. Объем буферов определяется номером ступени (на каждой следующей ступени осуществляется децимация с коэффициентом 2 и, следовательно, объем данных уменьшается в два раза). Число буферов зависит от количества коэффициентов цифровых фильтров. В рассматриваемой реализации использовались фильтры с 30 коэффициентами. Число буферов должно быть достаточным для того, чтобы хранить предисторию (предыдущие отсчеты входного сигнала) для гладкого сопряжения соседних фреймов входного сигнала. Поэтому с нулевой по третью ступень число буферов равно двум, для четвертой ступени – трем, для пятой ступени – пяти, для шестой – девяти, для седьмой ступени – шестнадцати и для последней ступени – тридцати.

Переключение буфера для записи данных происходит после его заполнения по команде запуска процедуры анализа, поступающей от DSP. Переключение буфера для чтения данных происходит после завершения цикла анализа текущего фрейма. Дисциплина работы с буферами – кольцевая как для процедуры записи, так и для процедуры чтения. Последний заполненный буфер является входным для вычислительного ядра ступени. Запись данных следующего фрейма осуществляется в следующий свободный буфер. Выходные данные предыдущей ступени пишутся в свободный входной буфер (буфер для записи) следующей ступени.

Необходимость наличия в памяти трех адресных портов объясняется наличием трех независимых источников адреса ячеек памяти:

- адресация при записи в блок памяти;
- адресация для чтения данных в вычислительное ядро ступени;
- адресация при чтении данных ступени на шину данных и далее в DSP.

Блок памяти нулевой ступени имеет дополнительный выход через буфер на шину данных системы.

Адресация при записи организуется с помощью счетчика адреса записи. При чтении данных на шину данных адресацию организует блок сопряжения с DSP с помощью шины `adr_mem`. При чтении данных в вычислительное ядро ступени счетчик стартового адреса обеспечивает хранение и наращивание начального адреса для получения каждого выходного отсчета, счетчик адреса чтения – перебор соседних входных отсчетов. С помощью мультиплексора начальных условий (на структурной схеме - MUX) обеспечивается подключение к адресному входу блока памяти либо адреса, формируемого счетчиком адреса чтения, либо адреса отсчетов предистории, хранящихся в других буферах, записанных ранее. В последнем случае требуемый адрес поступает от схемы формирования адреса предистории. Сигнал, управляющий мультиплексором, формируется в вычислительном ядре ступени.

В состав блока анализа также входит память коэффициентов фильтров и счетчик для адресации этой памяти. Коэффициенты фильтров подаются на вычислительное ядро каждой ступени параллельно.

DCM (Digital Clock Manager) формирует из тактовой частоты 50 МГц сигнал удвоенной частоты синхронизации `clkx2` и сигнал `clk_log`, которые необходимы для управления блоками памяти ступеней.

20.5.2 Структура блока сопряжения с DSP

Блок сопряжения с DSP предназначен для приема команд от DSP, сопряжения анализатора с шиной данных и адреса, формирования сигнала прерывания в DSP, формирования сигнала запуска анализатора, формирования адреса при чтении содержимого блоков памяти ступеней, формирования сигналов управления выходными буферами шины данных.

Структура блока сопряжения приведена на рис. 20.15. Шина адреса, сигналы `cs`, `wr`, `rd` от DSP подключены на вход дешифратора команд. На выходе дешифратора команд после формирователей вырабатываются управляющие сигналы (команды) для инициализации операций обмена с процессором или запуска анализатора. Формирователи обеспечивают длительность управляющих сигналов равной длительности одного такта частоты синхронизации `clk` (20 нс).

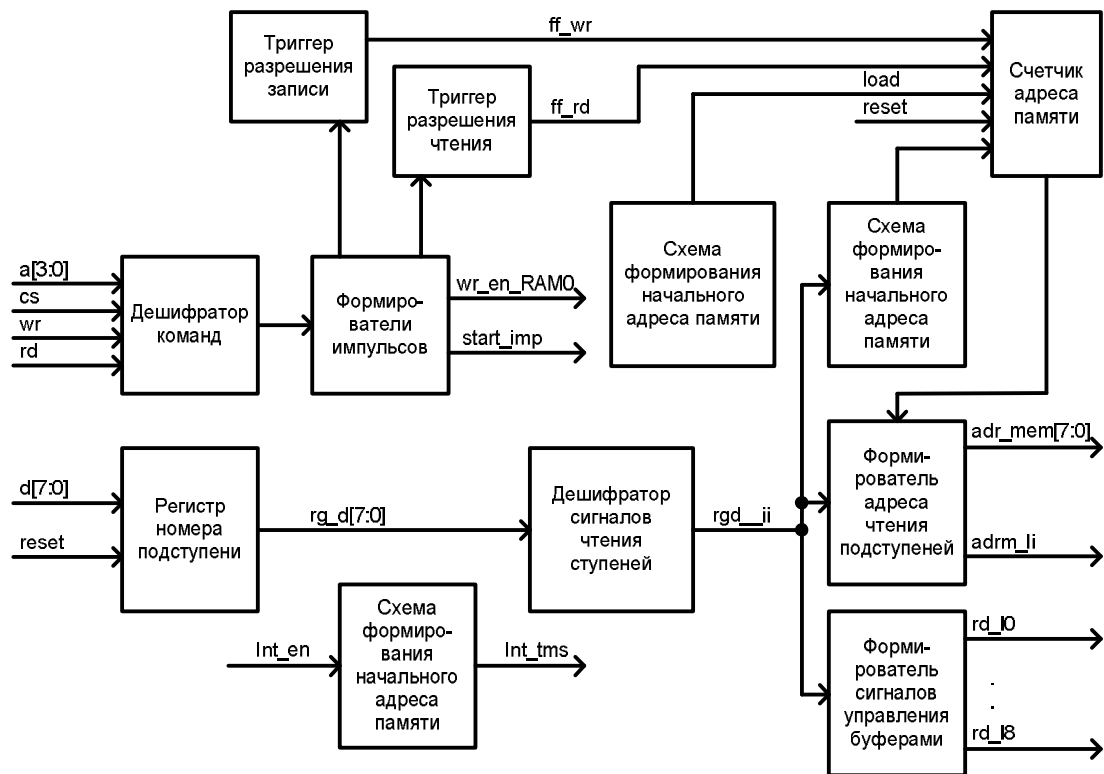


Рис. 20.15 – Структура блока сопряжения с DSP

При обнаружении команды записи блока (код 3) устанавливается триггер ff_wr, который разрешает работу счетчика адреса памяти st_adr для последующей записи 256 слов данных в блок памяти нулевой ступени. Непосредственная запись слов производится командой с кодом 4. Эта команда сопровождает каждое слово данных на шине данных и наращивает счетчик адреса на единицу. Триггер разрешения записи сбрасывается после приема 256 слов.

При обнаружении команды чтения блока (код 5) устанавливается триггер разрешения чтения блока ff_rd. Установка этого триггера также разрешает работу счетчика адреса. Одновременно по шине данных в блок сопряжения с DSP поступает номер подступени, из которой необходимо считать данные. Этот номер запоминается в регистре номера подступени и дешифрируется. На основе дешифрованной комбинации формируется адрес памяти для чтения из блока памяти соответствующей подступени и сигналы управления выходными буферами с тремя устойчивыми состояниями для подключения выходных линий данных блоков памяти подступеней к шине данных процессора (rd_i0 – rd_i8).

Непосредственное чтение слов из соответствующего блока памяти организуется командой с кодом 6. При этом наращивается содержимое счетчика адреса памяти и после передачи требуемого количества слов обмен данными с процессором прекращается.

Сигнал прерывания int_tms формируется после завершения обработки данных в вычислительных ядрах ступеней.

20.5.3 Вычислительное ядро

Каждое вычислительное ядро (рис. 20.16) содержит схему управления мультиплексором предистории и процессорный блок. Процессорный блок осуществляет цифровую фильтрацию входного сигнала. Процессорные блоки всех ступеней идентичны.

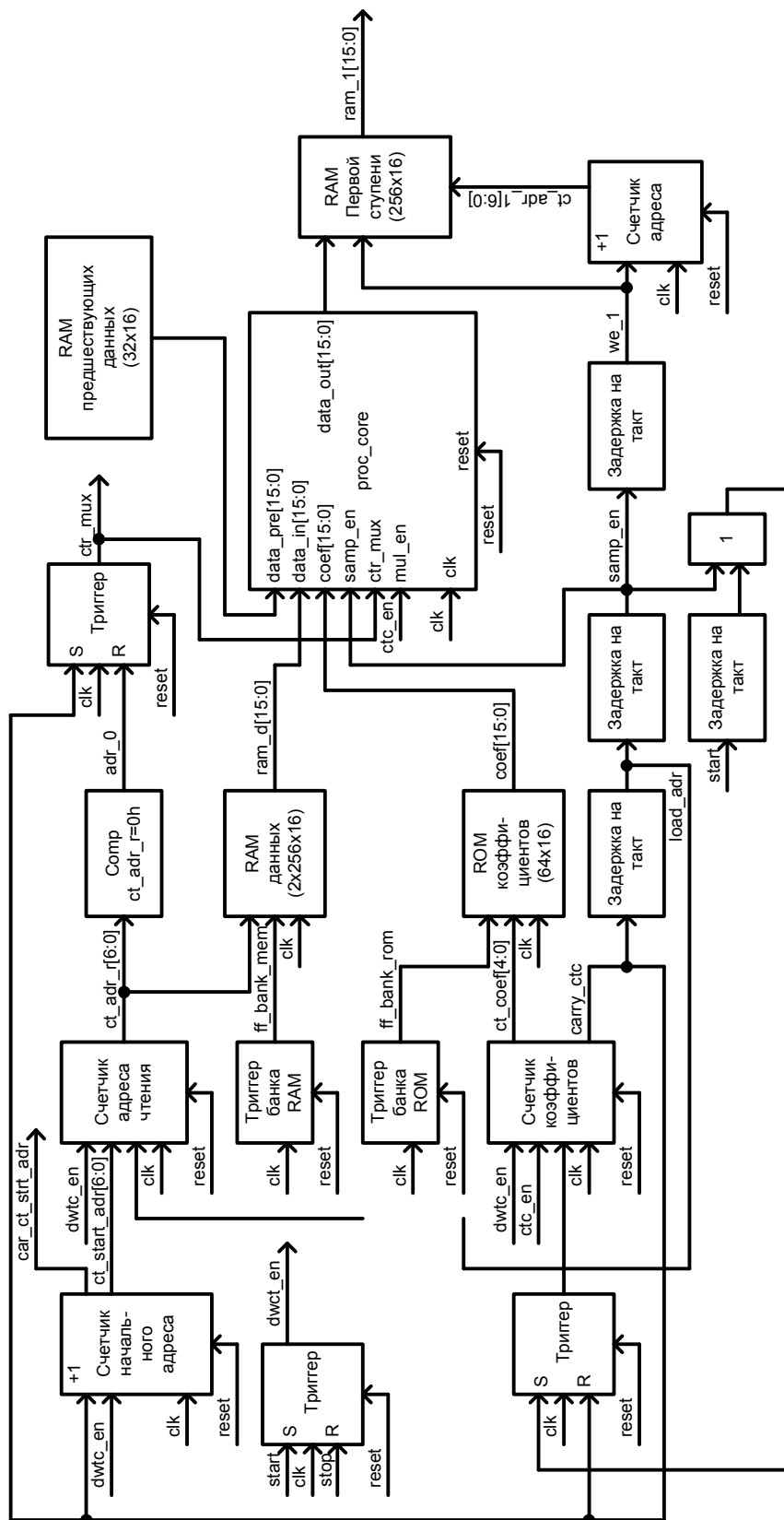


Рис. 20.16 – Структура вычислительного ядра

Оценки времени обработки аудиосигнала показали, что при тактовой частоте вычислительного ядра 50 МГц цифровую фильтрацию можно успевать производить в реальном масштабе времени, последовательно обрабатывая каждый входной фрейм фильтром нижних частот, а затем фильтром верхних частот на одном процессорном ядре с MAC структурой (структу-

рой умножитель-аккумулятор). Такой подход позволяет рационально использовать ресурсы ПЛИС. Структура процессорного блока представлена на рис. 20.17.

Входная шина данных процессорного блока подключена к одному входу блочного умножителя XC3S200, на второй вход умножителя подаются коэффициенты фильтров. Результат умножения запоминается в регистре `rg_mul`. Взвешенные коэффициентами фильтра входные отсчеты накапливаются в аккумуляторе `acc`, имеющем удвоенную по сравнению с входным словом разрядность. Для синхронизации работы умножителя и аккумулятора используется триггер `acc_en`. Сигнал `samp_en` формируется по завершении перебора коэффициентов фильтра и предназначен для сброса аккумулятора и перезаписи полученного выходного отсчета в выходной регистр.

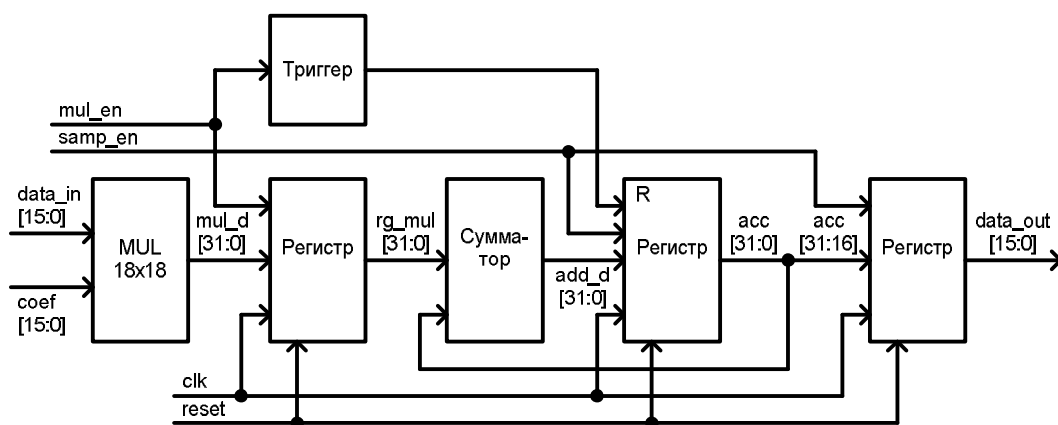


Рис. 20.17 – Структура процессорного блока

20.5.4 Структура памяти ступени анализатора

Блок памяти каждой ступени реализован на базе блочной памяти семейства FPGA Spartan III. С учетом разрядности обрабатываемых данных блок памяти сконфигурирован как двухпортовый с организацией 1024x18, причем используется основная память 1024x16, а память паритета 1024x2 не используется. Поскольку максимальный размер буфера не превышает 256x16 и в каждой ступени не менее двух буферов, то один блок памяти FPGA можно использовать для двух блоков памяти ступеней, при этом каждый блок памяти ступеней должен иметь три адресных порта. Для возможности параллельной работы двух ступеней с таким блоком памяти блочная память FPGA работает на удвоенной частоте синхронизации (100 МГц), а адресация памяти разных блоков осуществляется старшим разрядом адреса. Структура двойного блока памяти (для двух ступеней анализа) приведена на рис. 20.18.

Один порт (порт а) двухпортовой блочной памяти `ramb16_s18_s18` используется для работы с одной ступенью, второй порт (порт b) – для работы со второй ступенью. Поскольку обращение по этим двум портам происходит к разным областям памяти, то конфликтов при параллельной работе с двумя ступенями не возникает.

Три источника адреса коммутируются на адресный вход каждого порта с помощью двух ступеней мультиплекторов. Для разделения во времени адресного входа каждого порта используется сигнал `clk_log`, формируемый блоком DCM. Этот сигнал совпадает с сигналом `clk`, но в отличие от последнего, он доступен для использования логикой ячеек FPGA. Таким обра-

зом в течение одного полупериода сигнала clk (10 нс) с памятью выполняются операции записи, в течение другого полупериода – операции чтения. Дальнейшее разделение во времени двух источников адреса при чтении памяти реализуется с помощью сигнала $dwtc_en$. Этот сигнал устанавливается после запуска процессором очередного цикла анализа и сбрасывается после его завершения. Поэтому при установленном сигнале $dwtc_en$ осуществляются операции чтения в вычислительное ядро системы, а при сброшенном – операции чтения из блока памяти на шину данных. Поскольку операции чтения на шину данных могут осуществляться только после завершения анализа, то конфликтов адресации не происходит.

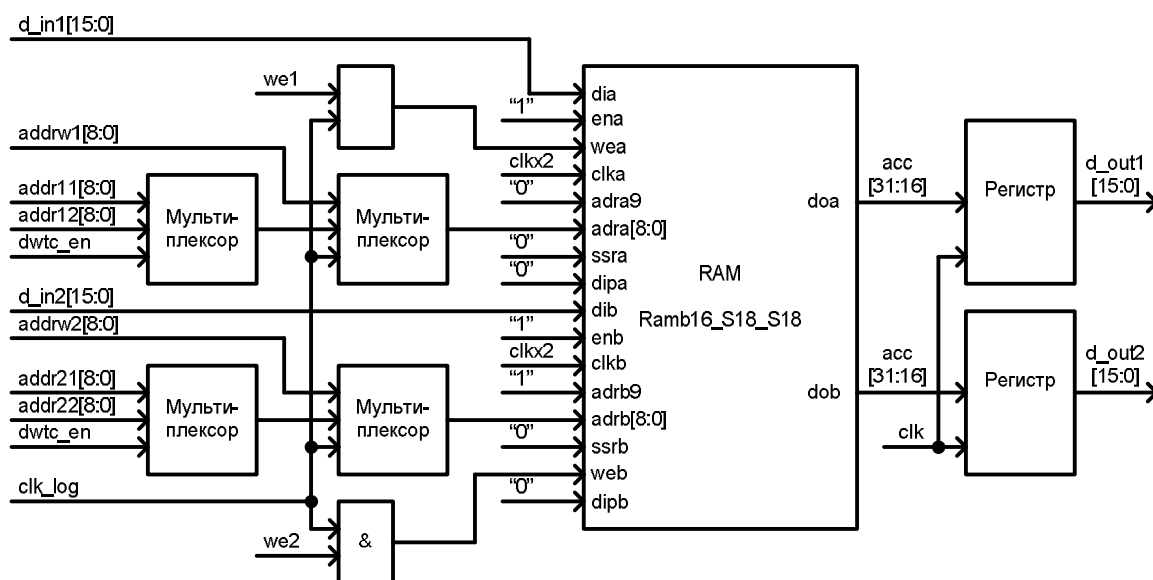


Рис. 20.18 – Структура блока памяти

Если требуется параллельная запись и чтение в каждый блок памяти с выводом выходных данных, то необходимо дополнить приведенную схему выходными регистрами, запоминающими выходные данные памяти по инверсному сигналу clk . Полученная пара регистров для каждого выхода двухпортовой блочной памяти будет содержать отдельно данные при операциях чтения и операциях записи.

С учетом наличия на входе блока памяти мультиплексора начальных условий критическим путем проекта является путь от выхода счетчика адреса чтения до адресного входа блочной памяти. Задержка распространения сигнала по этому пути определяет предельную тактовую частоту работы анализатора. Моделирование работы анализатора в режиме post place and route показало, что проект будет работоспособен до тактовой частоты 70 МГц.

20.6 Заключение

Из табл. 20.4 следует, что для реализации анализатора использовано около 40% кристалла по логическим блокам, 50% блочной памяти, 66% матричных умножителей. Оставшихся ресурсов кристалла FPGA достаточно для реализации на этом же кристалле устройства, осуществляющего синтез. Для экономии числа матричных умножителей можно использовать их раз-

деление во времени для двух ступеней, аналогично решению, примененному для построения блоков памяти.

Табл. 20.4 – Результат использования устройства

Использование логических элементов	Использовано	Доступно	Израсходовано
Число Flip Flops	965	3 840	25%
Число LUT с 4-мя входами	875	3 840	22%
Распределение логики			
Число занятых КЛБ	784	1 920	40%
Число КЛБ только для логики	784	784	100%
Общее число LUT с 4-мя входами	913	3 840	23%
Блоков IO	26	173	15%
Блоков RAMs	6	12	50%
MULT18x18s	8	12	66%
GCLKs	2	8	25%
DCM	1	4	25%

20.7 Список использованных источников

- Spanias A., Painter T., Atti V., Audio signal processing and coding, Hoboken, NJ: John Wiley & Sons, Inc., 2007, 464 p.
- Jayant N., Noll P., Digital coding of waveforms: principles and applications to speech and video, Englewood Cliffs, NJ: Prentice-Hall, 1984, 964 p.
- Sinha D., Tewfik A., Low bit rate transparent audio compression using adapted wavelets. IEEE Trans. on Signal Processing, Dec. 1993, Vol. 41, pp. 3463-3479.
- Vetterli M., Kovacevic J., Wavelets and subband coding, Englewood Cliffs, NJ, Prentice-Hall, 1995, 488 p.
- Wickerhauser M.V., Adaptive Wavelet Analysis from Theory to Software, A.K. Peters Ltd., Massachusetts, 1994, 486 p.
- Zwicker E., Fastl H., Psychoacoustics: facts and models, Berlin, Springer-Verlag, 1990, 380 p.
- Petrovsky Al., Krahe D., Petrovsky A.A., Real-time wavelet packet-based low bit rate audio coding on a dynamic reconfigurable system, AES, Convention paper 5778, 114th Convention, 2003 March 22-25, Amsterdam, The Netherlands, 22 pages.
- Wu X., Li X., Chen H., Programmable wavelet packet transform processor, Electronics letters, 18th March, 1999, Vol. 35, No. 6, pp. 449-450.
- Petrovsky Al., Petrovsky A., Reconfigurable processing: the solution to audio coding in real-time, Proc. of 9th Int. Conf. "Mixed design of Integrated circuits and systems"(MIXDES'2002), June, Wroclaw, Poland, 2002, pp. 629-634.