

Министерство образования Республики Беларусь  
Учреждение образования  
«Белорусский государственный университет  
информатики и радиоэлектроники»

Факультет компьютерных систем и сетей

Кафедра электронных вычислительных средств

## ***СИСТЕМЫ ОБРАБОТКИ МЕДИАДАННЫХ***

*Рекомендовано УМО по образованию в области информатики  
и радиоэлектроники для специальности 1-40 80 01  
«Элементы и устройства вычислительной техники и систем управления»  
в качестве пособия*

Минск БГУИР 2018

УДК 004.383.3:621.391(076)  
ББК 32.811.3я73+32.971.32-04я73  
С40

**Авторы:**

А. А. Петровский, М. И. Вашкевич, И. С. Азаров, Н. А. Петровский,  
В. В. Ключеня, Д. С. Лихачёв

**Рецензенты:**

кафедра радиофизики и цифровых медиатехнологий Белорусского  
государственного университета (протокол №3 от 01.11.2016);

главный научный сотрудник государственного научного учреждения  
«Объединенный институт проблем информатики Национальной академии  
наук Беларуси», доктор технических наук, профессор Г. И. Алексеев

**Системы** обработки медиаданных : пособие / А. А. Петровский  
С40 [и др.]. – Минск : БГУИР, 2018. – 120 с. : ил.  
ISBN 978-985-543-351-5.

Представлены современные методы и алгоритмы систем обработки медиаданных: речевых и звуковых сигналов, а также новые трансформационные преобразования для анализа и компрессии цифровых изображений по схеме L2L. Показан ряд примеров практических приложений для систем мультимедиа: изменение частоты основного тона речевого сигнала; коррекция певческого голоса; параметрический аудиоречевого кодер; преобразование цифровых изображений на основе параунитарных банков фильтров в алгебре кватернионов и блочной лестничной структуре прямого и обратного дискретного косинусного преобразования.

Может быть использовано при изучении дисциплин «Цифровая обработка данных», «Системы обработки медиаданных».

**УДК 004.383.3:621.391(076)**  
**ББК 32.811.3я73+32.971.32-04я73**

**ISBN 978-985-543-351-5**

© УО «Белорусский государственный  
университет информатики  
и радиоэлектроники», 2018

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ .....	5
<b>ЧАСТЬ 1 ПАРАМЕТРИЧЕСКИЙ АНАЛИЗ РЕЧЕВЫХ И ЗВУКОВЫХ СИГНАЛОВ</b>	
1 СИНУСОИДАЛЬНОЕ ПРЕДСТАВЛЕНИЕ ЗВУКОВЫХ И РЕЧЕВЫХ СИГНАЛОВ ДЛЯ ОБРАБОТКИ В СИСТЕМАХ МУЛЬТИМЕДИА .....	7
1.1 Синусоидальное и гармоническое представления сигнала .....	7
1.2 Гармонический анализатор и синтезатор .....	10
1.3 Синтез мгновенного гармонического преобразования .....	11
1.4 Синтез мгновенного гармонического преобразования с частотно-временным масштабированием .....	16
1.5 Оценка мгновенных синусоидальных параметров для обработки звуковых сигналов .....	19
2 СИСТЕМЫ ПАРАМЕТРИЧЕСКОЙ ОБРАБОТКИ РЕЧИ ДЛЯ ПРАКТИЧЕСКИХ ПРИЛОЖЕНИЙ .....	22
2.1 Параметрический анализ и синтез речевого сигнала на основе стационарной периодической модели .....	22
2.2 Система гибридного моделирования речевого сигнала с изменяющимися синусоидальными параметрами .....	26
2.3 Реализация системы изменения частоты основного тона в реальном масштабе времени .....	35
2.4 Результаты экспериментов изменения частоты основного тона речевого сигнала .....	40
2.5 Система коррекции певческого голоса .....	44
3 ПАРАМЕТРИЧЕСКИЙ АНАЛИЗ АУДИОСИГНАЛОВ НА ОСНОВЕ РАЗРЕЖЕННОЙ АППРОКСИМАЦИИ С ПЕРЦЕПТУАЛЬНО-ОПТИМИЗИРОВАННЫМ ПДВП СЛОВАРЕМ ВЕЙВЛЕТ-КОЭФФИЦИЕНТОВ .....	48
3.1 Разреженная аппроксимация аудиосигналов. Постановка задачи .....	48
3.2 Параметрический анализ аудиосигналов на основе разреженной аппроксимации и полным ПДВП-словарем вейвлет-коэффициентов .....	50
3.3 Разреженная аппроксимация аудиосигналов с перцептуально-оптимизированным ПДВП-словарем вейвлет-коэффициентов .....	54
3.4 Параметрический аудиоречевой кодер на основе разреженной аппроксимации с перцептуально-оптимизированным словарем частотно-временных функций .....	65
<b>ЧАСТЬ 2 СОВРЕМЕННЫЕ ПРЕОБРАЗОВАНИЯ ДЛЯ ПРИЛОЖЕНИЙ КОДИРОВАНИЯ ИЗОБРАЖЕНИЙ</b>	
4 ТРАНСФОРМАЦИОННОЕ КОДИРОВАНИЕ ЦИФРОВЫХ ИЗОБРАЖЕНИЙ .....	68

4.1	Система трансформационного кодирования .....	68
4.2	Необходимые свойства декоррелирующего преобразования в системе кодирования изображений .....	73
4.3	Мотивация построения кодера изображений по схеме L2L: lossless-to lossy.....	76
4.4	Структурный синтез преобразователей для систем трансформационного кодирования изображений .....	78
4.5	Решетчатые структуры регулярного ПУБФ на основе алгебры кватернионов .....	86
5	АЛГОРИТМЫ ВЫЧИСЛЕНИЯ ДКП.....	94
5.1	Применение ДКП в компрессии изображений.....	94
5.2	Основные типы и свойства ДКП.....	96
5.3	Вычислительная сложность алгоритмов ДКП .....	99
5.4	Алгоритмы вычисления ДКП.....	100
5.5	Преобразователь ДКП – ОДКП на основе блочной лестничной структурной параметризации .....	106
5.6	Преобразование ДКП – ОДКП с перфективной реконструкцией.....	109
6	ПРИМЕНЕНИЕ МЕТОДОВ ТРАНСФОРМАЦИОННОГО КОДИРОВАНИЯ ДЛЯ ОБРАБОТКИ ИЗОБРАЖЕНИЙ .....	113
6.1	Применение $Q$ -ПУБФ в компрессии изображений.....	113
6.2	Компрессия изображений на основе ДКП – ОДКП.....	115
	ЛИТЕРАТУРА.....	118

## ВВЕДЕНИЕ

Термин «мультимедиа» в настоящее время приобрел необычную популярность. Это объясняется тем, что появилась возможность связывания в единое целое визуальной, акустической и текстовой информации, т. е. сделан большой шаг от простых массовых продуктов, например, таких как сотовый телефон и MP3 проигрыватель, к амбициозным проектам – человеко-машинным интерфейсам. Стремительное внедрение цифровых технологий в такие приложения, как аудиокниги, системы радиовещания и телевидения, непрерывное увеличение передач через Интернет, информационное наполнение которых может быть различно и не ограничено только речью или музыкой, определило необходимость в объединенном аудиокодере с низкой скоростью битового потока, который одинаково хорошо работает со всеми типами звукового информационного наполнения. Общая философия перцептуального кодера взаимосвязана с выбором метода частотно-временного анализа – банка цифровых фильтров. Развитие этих работ в настоящее время идет по пути построения перцептуальных субполосных аудиокодеров.

В настоящее время, благодаря росту производительности вычислительных платформ, технические возможности значительно расширились и позволяют проектировать системы цифровой обработки речи, использующие более сложные методы моделирования. В связи с этим разработка способов описания речевого сигнала как непрерывного процесса нестационарными («мгновенными») параметрами приобретает высокую теоретическую и практическую ценность. Прикладное значение параметрического моделирования в сочетании с биоинспирированными принципами обработки повышается в контексте задач шумоподавления и повышения разборчивости речи.

Особую значимость приобретает проблема сжатия цифровых изображений на основе использования трансформационного преобразования. В настоящее время интерес представляют схемы компрессии изображений, работающие по принципу L2L (lossless-to-lossy) и осуществляющие сжатие и восстановление цифровых изображений как без потерь (режим lossless, трансформационное преобразование обратимо и реализуется в целочисленной арифметике), так и с контролируемым внесением артефактов (режим lossy). Кодирование по принципу L2L унифицирует схемы сжатия цифровых изображений в режимах lossy и lossless, когда требуется высокое качество и степень компрессии. Это обеспечивает возможность широкого применения схем кодирования изображений в соответствии с мультимедийными приложениями. Например, в стандартах JPEG и JPEG2000 режимы lossless и lossy не имеют между собой совместимости, так как трансформационные преобразования оптимизированы для работы в своем режиме. В стандарте JPEG XR поддерживается схема L2L, однако эффективность кодирования недостаточная, особенно для изображений, у которых много высокочастотных компонент.

Вышесказанное определило цель данного пособия – показать методы, алгоритмы и реализации анализаторов речевых и звуковых сигналов, которые базируются на теории цифровых банков фильтров и параметрического представления аудиосигнала в виде квазипериодической и стохастической составляющих, а также новые трансформационные преобразования для анализа и компрессии цифровых изображений по схеме L2L.

Библиотека БГУИР

# ЧАСТЬ 1 ПАРАМЕТРИЧЕСКИЙ АНАЛИЗ РЕЧЕВЫХ И ЗВУКОВЫХ СИГНАЛОВ

## 1 СИНУСОИДАЛЬНОЕ ПРЕДСТАВЛЕНИЕ ЗВУКОВЫХ И РЕЧЕВЫХ СИГНАЛОВ ДЛЯ ОБРАБОТКИ В СИСТЕМАХ МУЛЬТИМЕДИА

### 1.1 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 и трансформ-кодер, а также в кодировании широкополосных звуковых сигналов.

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. Синусоидальная модель неэффективна для описания шумовых и импульсных звуков, в связи с этим на практике часто используется гибридная модель, которая использует разные способы параметрического представления сигнала. В современных системах компрессии используют раздельное кодирование синусоидальной, шумовой и транзиентной (переходной) составляющих. Первая из составляющих 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 и применения специальных способов шумоподавления. Аналогичное гибридное параметрическое представление используется в системах обработки звука для создания специальных эффектов (например, изменение масштаба времени, основного тона, тембра и др.), а также в системах синтеза речи по тексту. Речевой сигнала также удобно описывать гибридной моделью, поскольку 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. 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.

Ключевым вопросом при использовании синусоидальной модели является определение гармонических параметров, так как от точности их оценки зависит результат применения всего подхода в целом. Существует большое количество способов для выполнения гармонического анализа и это научное направление постоянно совершенствуется. Одной из основных задач 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 изменений в самой системе кодирования.

Основным способом оценки гармонических параметров сигнала в настоящее время является дискретное преобразование Фурье (ДПФ). ДПФ представляет анализируемый сигнал в виде суммы периодических компонентов, которые имеют постоянную частоту и амплитуду на протяжении всего фрейма анализа. Такое представление делает способ ограниченно применимым в том случае, когда оцениваемые параметры являются нестационарными и должны рассматриваться как мгновенные.

Так как синусоидальное моделирование в большинстве случаев можно применить непосредственно к входу системы, входной сигнал удобно рассматривать как комбинацию периодической и остаточной компонент. Таким образом, сигнал  $s(n)$  можно записать в виде соотношения

$$s(n) = \sum_{k=1}^K \text{MAG}_k(n) \cos \varphi_k(n) + r(n), \quad (1.1)$$

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:

$$\varphi_k(n) = \sum_{i=0}^n \frac{2\pi f_k(i)}{F_s} + \varphi_k(0), \quad (1.2)$$

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:



$$f_k(n) = kf_0(n). \quad (1.3)$$

Гармоническая модель используется для кодирования речевых сигналов с высоким коэффициентом сжатия, так как обеспечивает чрезвычайно эффективное описание вокализованных фрагментов речи.

Задачей гармонического анализа является определение синусоидальных параметров  $MAG_k(n)$ ,  $f_k(n)$  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 создают переходную (транзиентную) компоненту;

- синусоидальные компоненты достаточно продолжительны, в противном случае они формируют либо переходную компоненту, 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, что накладывает определенные ограничения. В этом случае предполагается, что анализируемый сигнал является квазистационарным, т. е. 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 и кодирования речи, 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.

## 1.2 Гармонический анализатор и синтезатор

В системах мультимедиа, использующих синусоидальную или гармоническую модель, как правило, присутствуют: блок анализа (гармонический анализатор), определяющий гармонические параметры исходного сигнала, и блок синтеза (гармонический синтезатор), формирующий выходной сигнал из гармонических параметров.

Общая схема гармонического анализатора, используемого при обработке звука и речи, представлена на рисунке 1.1. Исходный сигнал разбивается на перекрывающиеся фреймы, которые преобразовываются из временной области в частотную при помощи какого-либо частотно-временного преобразования. Далее полученные спектральные представления фреймов сравниваются для нахождения синусоидальных компонент, удовлетворяющих заданным условиям (необходимая продолжительность, уровень амплитуды и т. д.). Параметрическое представление сигнала содержит наборы синусоидальных параметров (амплитуда, частота, фаза) для каждого фрейма. В гибридных анализаторах параметрическое представление также содержит сигнал-остаток в той или иной форме.

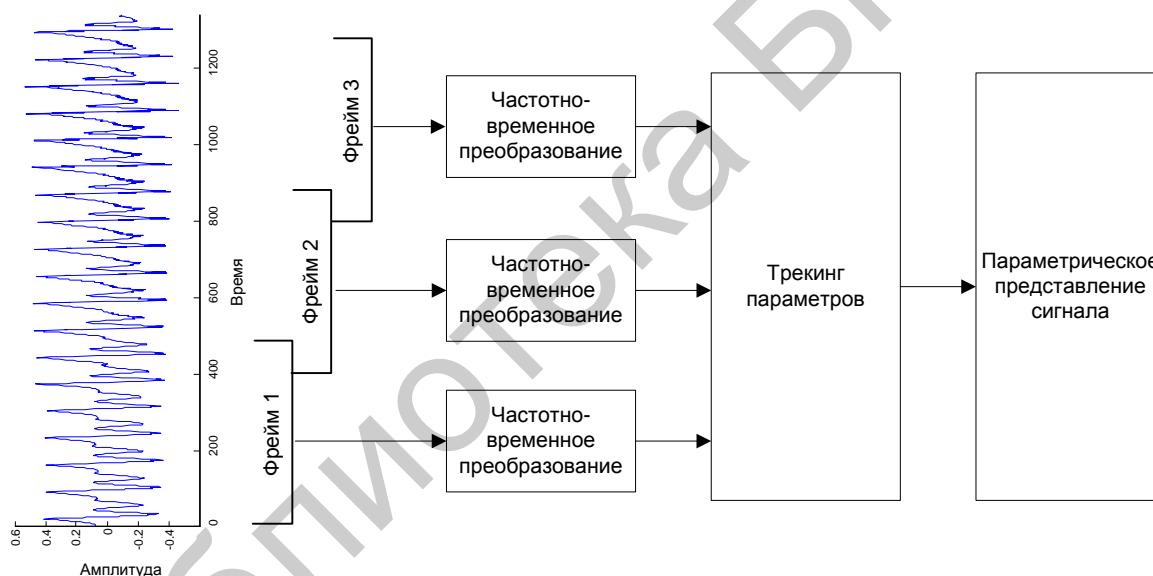


Рисунок 1.1 – Общая схема гармонического анализатора

Задача гармонического синтезатора заключается в получении сигнала во временной области из его параметрического представления. Для этого с целью согласования значения фазы на границах фреймов и соединения амплитудных огибающих без слышимых искажений применяется интерполяция синусоидальных параметров, после чего синтезируется выходной сигнал (рисунок 1.2).

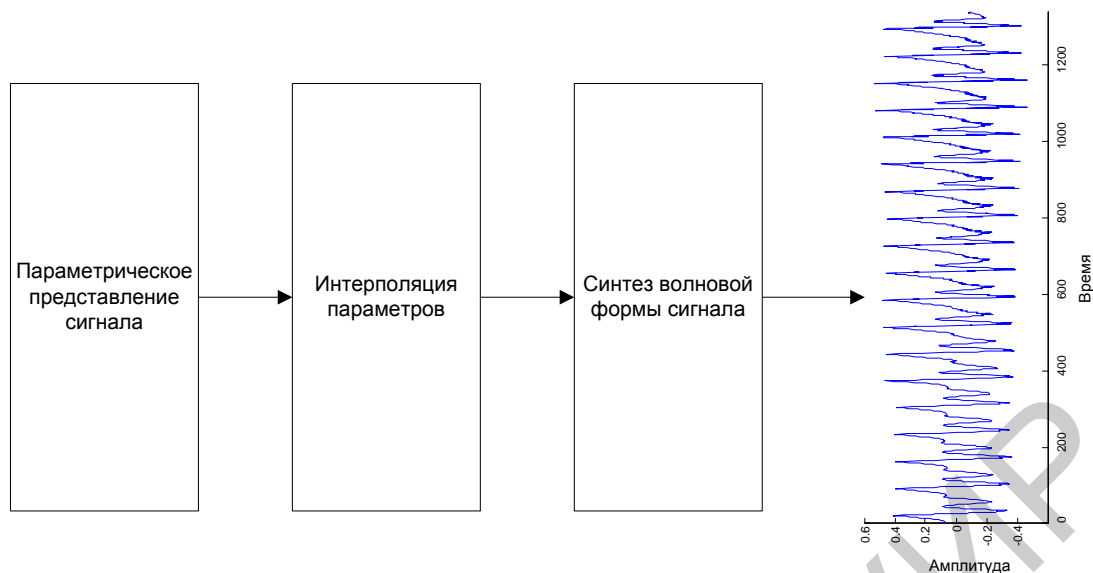


Рисунок 1.2 – Общая схема гармонического синтезатора

В гибридных системах выполняется отдельная обработка параметров каждой модели и используется отдельный синтез, после которого составляющие сигнала суммируются.

### 1.3 Синтез мгновенного гармонического преобразования

Для оценки мгновенных гармонических параметров можно использовать фильтр с комплексной импульсной характеристикой. На выходе фильтра формируется узкополосный комплексный сигнал, действительная и мнимая части которого находятся в квадратуре.

Для синтеза фильтра можно воспользоваться оконным методом. Пусть  $F_1$  и  $F_2$  – нормированные частоты из диапазона  $[0, \pi]$ , определяющие соответственно нижнюю и верхнюю границы полосы пропускания, тогда непрерывная импульсная характеристика искомого фильтра  $h(t)$  определяется следующим выражением:

$$\begin{aligned}
 h_{F_1, F_2}(t) &= \frac{1}{\pi} \int_0^{F_2} e^{-j\omega t} d\omega - \frac{1}{\pi} \int_0^{F_1} e^{-j\omega t} d\omega = \\
 &= \frac{e^{-j\omega t}}{-jt\pi} \Big|_0^{F_2} - \frac{e^{-j\omega t}}{-jt\pi} \Big|_0^{F_1} = \frac{e^{-jF_1 t} - e^{-jF_2 t}}{jt\pi}.
 \end{aligned} \tag{1.4}$$

Выразим импульсную характеристику через середину полосы пропускания  $F_c$  и половину ширины полосы  $F_\Delta$ , используя подстановку  $F_1 = F_c - F_\Delta$  и  $F_2 = F_c + F_\Delta$ :

$$\begin{aligned} h_{F_1, F_2}(t) &= \frac{e^{-jF_c t} e^{jF_\Delta t} - e^{-jF_c t} e^{-jF_\Delta t}}{jt\pi} = \frac{e^{-jF_c t} (e^{jF_\Delta t} - e^{-jF_\Delta t})}{jt\pi} = \\ &= 2 \frac{\sin(F_\Delta t)}{t\pi} e^{-jF_c t}. \end{aligned} \quad (1.5)$$

Сигнал на выходе фильтра представляет собой АМ- и ЧМ-функцию косинуса с ограниченным частотным диапазоном:

$$s_{F_1, F_2}(t) = s(t) \cdot h_{F_1, F_2}(t) = \text{MAG}(t) \cos(\varphi(t)), \quad (1.6)$$

с мгновенной амплитудой  $\text{MAG}(t)$ , фазой  $\varphi(t)$  и частотой  $f(t)$ , которые могут быть определены по следующим формулам:

$$\text{MAG}(t) = \sqrt{R^2(t) + I^2(t)}, \quad (1.7)$$

$$\varphi(t) = \arctan\left(\frac{-I(t)}{R(t)}\right), \quad (1.8)$$

$$f(t) = \varphi(t + 1) - \varphi(t), \quad (1.9)$$

где  $R(t)$  и  $I(t)$  действительная и мнимая части комплексного сигнала  $S_{F_1, F_2}(t)$  соответственно.

Можно показать, что при разбиении всего частотного диапазона сигнала на полосы произвольными частотами  $F_1, F_2, \dots, F_L$ , такими, что  $F_1 = 0; F_L = \pi; F_1 \leq F_2 \leq \dots \leq F_L$ , исходный сигнал  $s(t)$  равняется сумме действительных частей выходов фильтров анализа:

$$s(t) = \sum_{k=1}^{L-1} \text{real}[s_{F_k, F_{k+1}}(t)]. \quad (1.10)$$

Для этого представим сигнал  $s(t)$  в виде его свертки с импульсной характеристикой идеального фильтра низких частот  $h(t)$ :

$$\begin{aligned}
s(t) &= s(t) \cdot h(t) = s(t) \cdot \frac{\sin(t\pi)}{t\pi} = s(t) \cdot \int_{-\pi}^{\pi} \cos(t\omega) d\omega = \\
&= s(t) \cdot \left[ 2 \int_0^{\pi} \cos(\omega t) d\omega \right] = s(t) \cdot \left[ \sum_{k=1}^{L-1} 2 \int_{F_k}^{F_{k+1}} \cos(\omega t) d\omega \right] = \\
&= \sum_{k=1}^{L-1} s(t) \cdot \text{real}[h_{F_k, F_{k+1}}(t)] = \sum_{k=1}^{L-1} \text{real}[s_{F_k, F_{k+1}}(t)]. \quad (1.11)
\end{aligned}$$

Сложение субполосных комплексных сигналов  $s_{F_k, F_{k+1}}(t)$  дает в результате комплексный сигнал Габора, соответствующий исходному сигналу  $s(t)$ , который вычисляется при помощи соотношения

$$z(t) = \sum_{k=1}^{L-1} s_{F_k, F_{k+1}}(t) = s(t) + jH[s(t)]. \quad (1.12)$$

Для получения импульсной характеристики фильтра анализа конечной длины следует использовать некоторую оконную функцию  $w(t)$ . Выражение

(1.5) принимает следующий вид:

$$h_{F_1, F_2}(t) = 2 \frac{\sin(F_{\Delta} t)}{t\pi} w(t) e^{-jF_c t}. \quad (1.13)$$

Заметим, что  $h_{F_1, F_2}(t)$  представляет собой произведение импульсной характеристики идеального фильтра низких частот, оконной функции и комплексной экспоненты, выполняющей частотный сдвиг на заданную частоту  $F_c$ . Используя (1.6) и (1.14), введем понятие мгновенного гармонического преобразования в момент времени  $t = 0$  следующим образом:

$$S(F_{\Delta}, F_c) = \int_{-\infty}^{+\infty} 2 \frac{\sin(F_{\Delta} t)}{t\pi} w(t) s(t) e^{-jF_c t} dt = \int_{-\infty}^{+\infty} h_{F_{\Delta}}(t) s(t) e^{-jF_c t} dt. \quad (1.14)$$

Из выражения видно, что для заданной ширины полосы пропускания  $F_{\Delta}$  мгновенное гармоническое преобразование представляет собой преобразование Фурье исходного сигнала  $s(t)$ , взвешенного импульсной характеристикой фильтра низких частот  $h_{F_{\Delta}}(t) = 2 \frac{\sin(F_{\Delta} t)}{t\pi} w(t)$ . Данная особенность имеет важное практическое значение, поскольку делает возможным вычисление (1.14) эффективным способом при помощи быстрого преобразования Фурье. Момент оценки

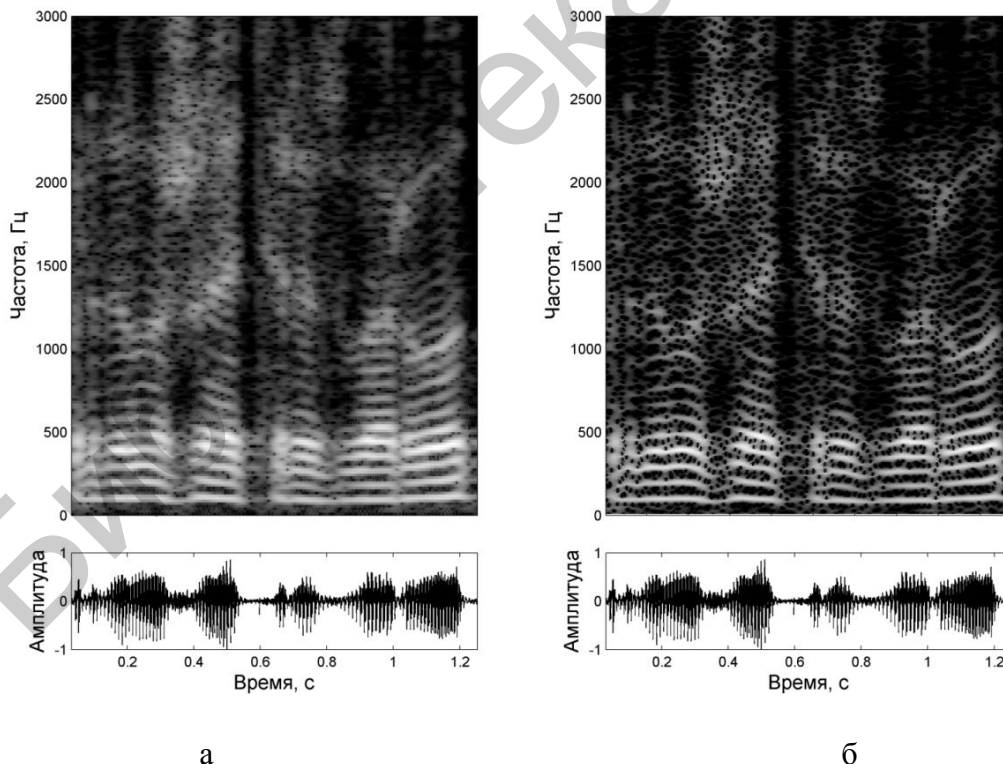
гармонических параметров определяется положением центра симметрии  $h_{F_\Delta}(t)$  относительно сигнала  $s(t)$  и поэтому в общем случае для оценки мгновенных гармонических параметров в произвольный момент  $t_0$  следует ввести смещение по времени:

$$S(F_\Delta, F_c, t_0) = \int_{-\infty}^{+\infty} h_{F_\Delta}(t - t_0) s(t) e^{-jF_c t} dt. \quad (1.15)$$

Мгновенное гармоническое преобразование для дискретного сигнала  $s(n)$  можно записать в следующем виде:

$$S(F_\Delta, F_c, n_0) = \sum_{n=0}^{N-1} h_{F_\Delta}(n - n_0) s(n) e^{-\frac{j2nF_c}{N}}. \quad (1.16)$$

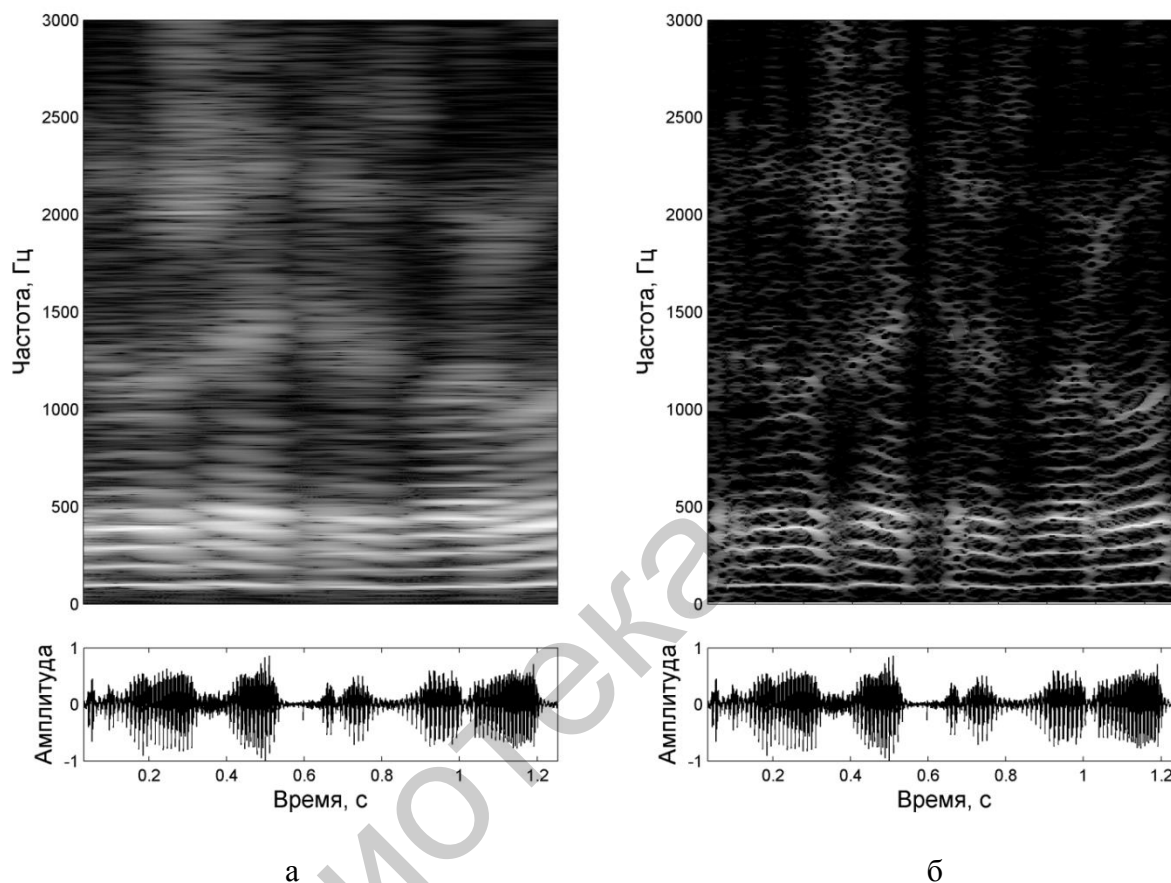
Ниже приведены спектрограммы речевого сигнала, полученные при помощи преобразования Фурье и мгновенного гармонического преобразования – рисунок 1.3 (размер окна анализа 64 мс). Увеличение длины окна анализа приводит к различным эффектам в спектре Фурье и мгновенном спектре – рисунок 1.4. При использовании более продолжительного окна (250 мс) спектр Фурье приобретает более высокое частотное разрешение, однако сильно усредняется по времени, что делает невозможным выполнение точной оценки периодических компонент нестационарного речевого сигнала.



а – преобразование Фурье; б – мгновенные параметры синусоидальной модели

Рисунок 1.3 – Оценка мгновенного спектра речевого сигнала (окно анализа 64 мс)

Спектральная оценка речевых гармоник, полученная при помощи мгновенного гармонического преобразования, также становится более точной по частоте и в то же время сохраняет высокое временное разрешение. Данное явление объясняется тем, что ширина полос пропускания не зависит от формата преобразования, но за счет удлинения импульсной характеристики осуществляется более точное разделение сигнала на субполосы.



а – преобразование Фурье; б – мгновенные параметры синусоидальной модели  
Рисунок 1.4 – Оценка мгновенного спектра речевого сигнала (окно анализа 250 мс)

Мгновенные гармонические параметры выходного сигнала могут быть рассчитаны в любой момент времени, принадлежащий анализируемому фрейму сигнала, причем этот момент не ограничивается дискретными отсчетами сигнала, так как выход фильтра записан в виде непрерывных функций. Очевидно, что полоса пропускания фильтра, задаваемая параметрами  $F_1$  и  $F_2$  (или  $F_c$  и  $F_\Delta$ ), должна содержать анализируемую компоненту и в случае наличия частотной модуляции полоса может оказаться достаточно большой (рисунок 1.5).

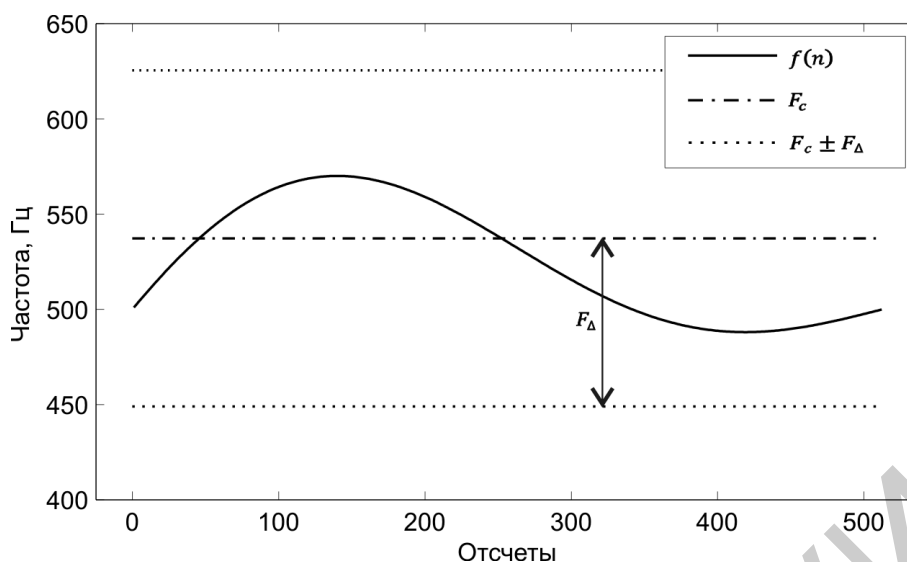


Рисунок 1.5 – Кратковременный гармонический анализ с использованием стационарного фильтра

Расширение полосы пропускания фильтра влияет на точность оценки мгновенной частоты. Анализируемая гармоническая компонента синтезирована с разрывом контура частоты, что дает возможность определить скорость реакции фильтров анализа.

В отдельных приложениях (вокализованная речь) для точного анализа частотно-модулированных компонент нужно использовать специальный фильтр, адаптирующийся к изменяющемуся частотному контуру.

#### 1.4 Синтез мгновенного гармонического преобразования с частотно-временным масштабированием

Для точной оценки синусоидальных параметров компонент с сильной частотной модуляцией должен быть использован частотно-модулированный фильтр, импульсная характеристика которого модулируется в соответствии с частотным контуром анализируемой компоненты. 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 (1.16) может быть использовано для определения мгновенного гармонического преобразования с частотным масштабированием:

$$S(F_\Delta, F_c(n), n_0) = \sum_{n=0}^{N-1} h_{F_\Delta}(n - n_0) s(n) e^{\frac{-j2n\varphi_c(n)}{N}}, \quad (1.17)$$

где  $\varphi_c(n) = \sum_{l=0}^n F_c(j)$ .



Мгновенные параметры (амплитуда, частота и фаза) вычисляются при помощи выражений (1.7) – (1.9). 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.

Таким образом, анализируемая частотно-модулированная компонента может быть выделена при помощи узкой масштабированной полосы пропускания (рисунок 1.6). Сужение полосы обеспечивает более точную локализацию и соответственно делает оценки параметров более точными. Это также позволяет применять мгновенный гармонический анализ к гармоникам высокого порядка речевого сигнала (рисунок 1.7). Чем выше номер гармоники, тем больше изменение ее частоты и импульсная характеристика ЧМ-фильтра изменяется соответствующим образом.

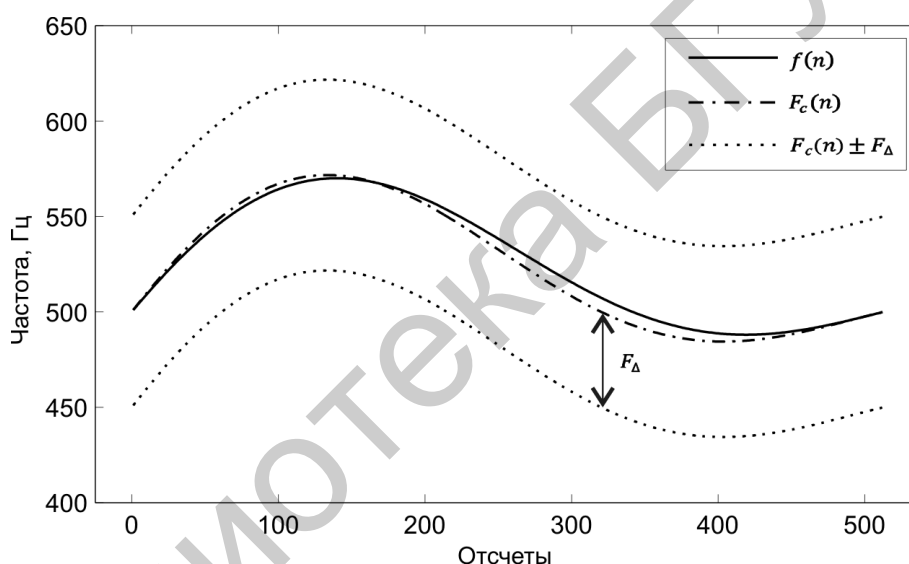
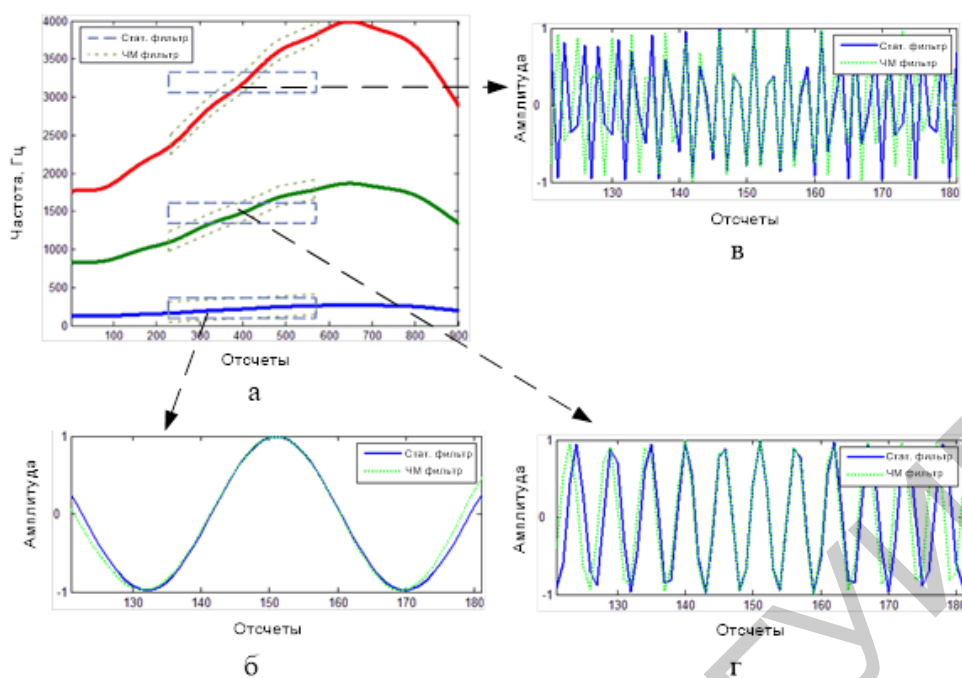


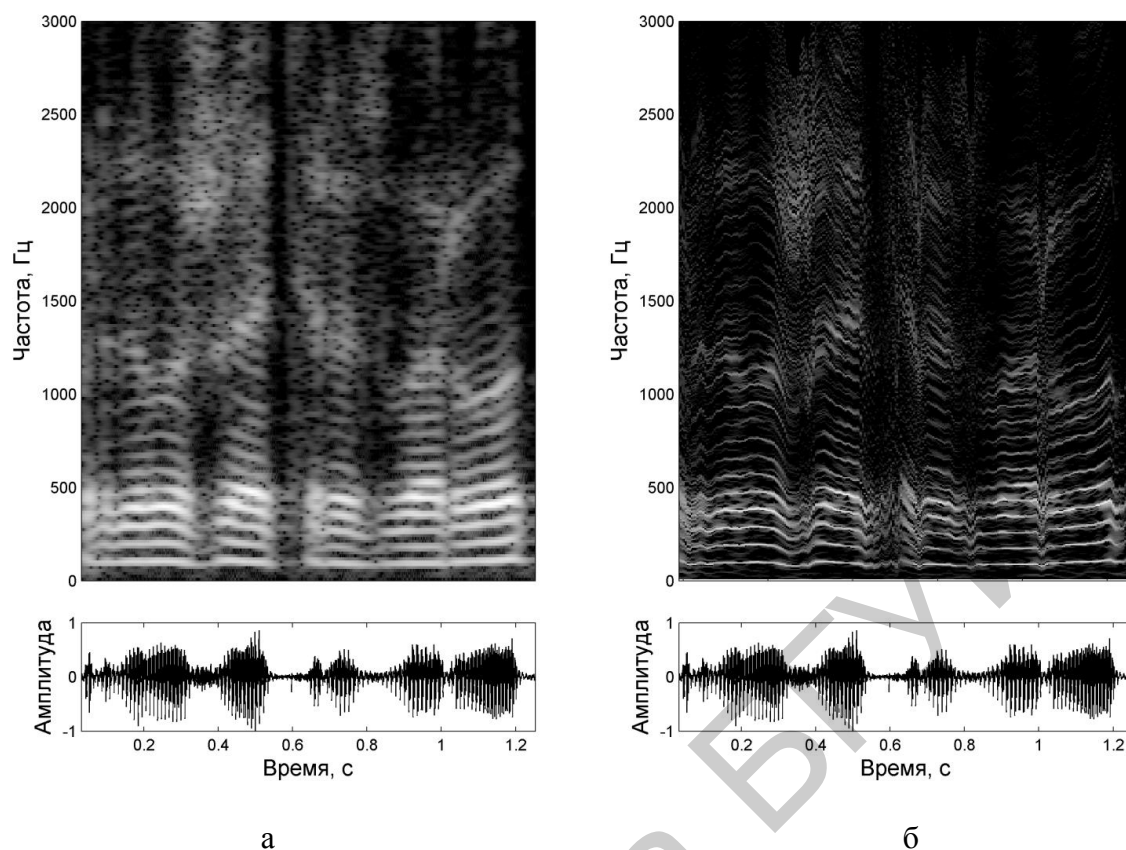
Рисунок 1.6 – Кратковременный гармонический анализ с использованием частотно-модулированного фильтра



а – полосы пропускания фильтров и частотные контуры гармоник;  
 б–г – импульсные характеристики фильтров для оценки основного тона, 7-й и 15-й гармоник соответственно

Рисунок 1.7 – Применение стационарного и ЧМ-фильтра для анализа вокализованной речи

Выполнение мгновенного гармонического преобразования, согласованного с контуром основного тона, существенно повышает частотно-временное разрешение оценок гармонических параметров речевого сигнала. При увеличении окна анализа можно получить еще более точные оценки параметров гармоник – рисунок 1.8.



а – преобразование Фурье (окно анализа 64 мс);  
 б – мгновенные параметры синусоидальной модели, полученные путем анализа преобразования, согласованного с контуром основного тона

Рисунок 1.8 – Оценка мгновенного спектра речевого сигнала при помощи гармонического преобразования, согласованного с контуром основного тона (окно анализа 250 мс)

### 1.5 Оценка мгновенных синусоидальных параметров для обработки звуковых сигналов

Для кодирования и обработки звуковых сигналов, как это было показано в подразделе 1.4, используются гибридные модели, разделяющие сигнал на периодическую, транзиентную и шумовую компоненты. Первичное разделение синусоидальной (периодической) части и остатка может выполняться путем гармонического анализа на основе фильтров. На рисунке 1.9 показана структура такой системы анализа, использующая гибридную модель параметризации.

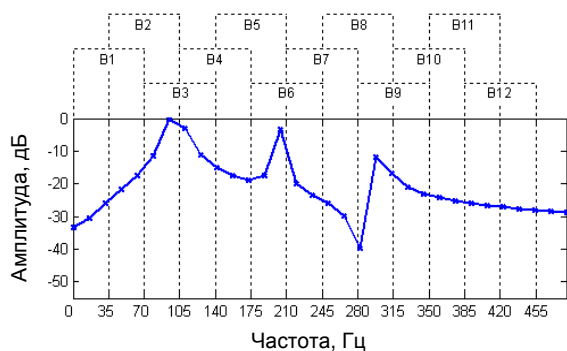


Рисунок 1.9 – Система гибридного анализа для кодирования звука и речи

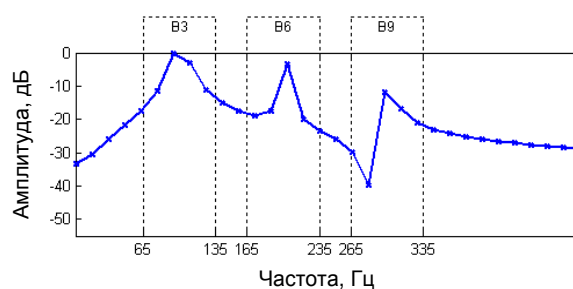
Вначале синусоидальный анализатор выполняет оценку параметров с постобработкой для выделения продолжительных периодических компонент. Затем выделенная периодическая часть анализируемого сигнала синтезируется синусоидальным синтезатором и вычитается из исходного сигнала. Остаток попадает на вход детектора транзиентов, который выполняет его дополнительное разделение.

Для локализации периодических компонент в частотной области 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 оцениваемой компоненты в центре полосы (рисунок 1.10).

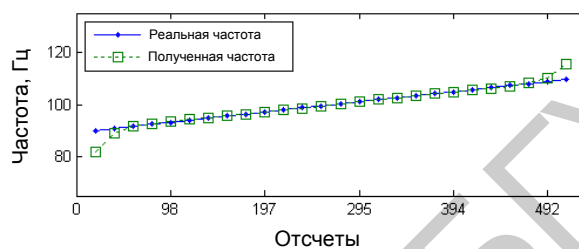
На первом шаге весь частотный диапазон анализируемого сигнала разбивается на полосы  $B_1, \dots, B_h$  ( $h$  – число полос) с центральными частотами  $F_C^{B_1}, \dots, F_C^{B_h}$  соответственно. На каждом шаге соответствующие оценочные мгновенные частоты  $f^{B_1}(n_c), \dots, f^{B_h}(n_c)$  определяются в момент времени, соответствующий центральному отсчету окна анализа  $n_c$ . Центральные частоты фильтров изменяются в соответствии с ними:  $F_C^{B_x} = f^{B_x}(n_c)$ .



а



б



в

а – первоначальное разбиение частотного диапазона; б – локализация компонент после второй итерации; в – оценка мгновенной частоты в полосе  $B_3$

Рисунок 1.10 – Кратковременный анализ звукового сигнала с итеративной локализацией периодических компонент,  $B_1$ – $B_{12}$  – полосы пропускания фильтров

После того как все энергетические максимумы локализованы, можно определить требуемые синусоидальные параметры (амплитудная огибающая, мгновенная частота и фаза). Параметры определяются при помощи выражений (1.7) – (1.9). Так как в процессе анализа несколько фильтров могут локализовать одну и ту же компоненту, с целью исключения дублирования параметров выполняется сравнение центральных частот  $F_C^{B_1}, \dots, F_C^{B_n}$ .

Чтобы отбросить короткие и неустойчивые компоненты, которые, очевидно, являются составляющими транзиентов или шума, выполняется слежение за полученными синусоидальными параметрами. Для выделения только достаточно продолжительных и стабильных значений сравниваются частоты и амплитуды компонент соседних фреймов.

## 2 СИСТЕМЫ ПАРАМЕТРИЧЕСКОЙ ОБРАБОТКИ РЕЧИ ДЛЯ ПРАКТИЧЕСКИХ ПРИЛОЖЕНИЙ

### 2.1 Параметрический анализ и синтез речевого сигнала на основе стационарной периодической модели

Одной из наиболее используемых в современных задачах мультимедиа системой параметрического моделирования (СПМ) речевых сигналов является TANDEM-STRAIGHT, основанная на раздельном моделировании периодической и стохастической составляющих сигнала. Система позволяет выполнять высококачественный синтез сигнала с измененными параметрами, что обуславливает ее использование во многих исследовательских и практических приложениях, таких, как конверсия голоса, синтез речи и певческого голоса, коррекция акцента, повышение привлекательности голоса, изменение просодики – темпа произношения и частоты основного тона.

Процедура анализа основывается на простом методе получения устойчивой оценки кратковременной спектральной плотности мощности путем усреднения двух оценок СПМ, вычисленных со смещением друг относительно друга на половину периода основного тона  $\frac{1}{2f_0}$ . Анализ выполняется согласованно с текущей частотой основного тона, причем окно анализа масштабируется таким образом, чтобы содержать фиксированное число периодов основного тона. Для получения гладкой огибающей амплитудного спектра оценка СПМ сглаживается путем фильтрации в области кепстральных коэффициентов.

На рисунке 2.1 показана схема анализа, используемая в TANDEM-STRAIGHT.

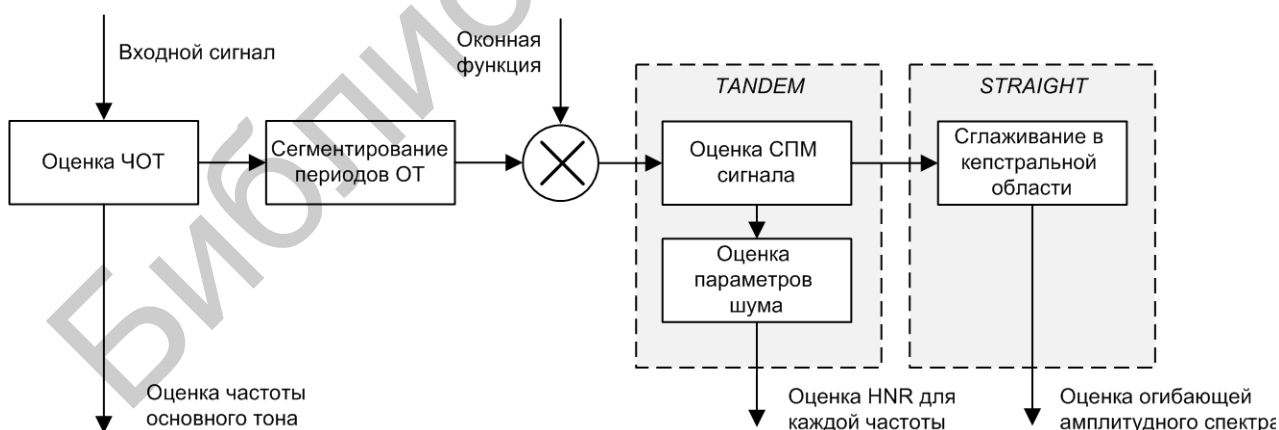


Рисунок 2.1 – Схема анализа речевого сигнала, используемая в системе параметрической обработки речи TANDEM-STRAIGHT

На первом шаге выполняется оценка частоты основного тона при помощи специального алгоритма с низким уровнем ошибок. На основе полученной

оценки частоты основного тона вычисляются оценки СПМ при помощи следующих выражений:

$$S(\omega, t) = \int s(\tau)w(\tau - t)e^{-j\omega\tau}d\tau, \quad (2.1)$$

$$P_T(\omega, t) = \frac{|S(\omega, t - \frac{t_0}{4})|^2 + |S(\omega, t + \frac{t_0}{4})|^2}{2}, \quad (2.2)$$

где  $S(\omega, t)$  – кратковременный спектр сигнала;

$P_T(\omega, t)$  – усредненная оценка СПМ из двух интервалов, смещенных друг относительно друга на половину длины периода основного тона;

$s(t)$  – анализируемый сигнал,

$w(t)$  – оконная функция;

$t_0$  – длина периода основного тона,  $t_0 = \frac{1}{f_0}$ .

Случайные колебания СПМ  $|S(\omega, t)|^2$  существенно ослабляются в оценке  $P_T(\omega, t)$  за счет усреднения. Сглаженная огибающая амплитудного спектра вычисляется при помощи следующих выражений:

$$C(\omega, t) = \int_{\omega_L}^{\omega} P_T(\lambda, t)d\lambda, \quad (2.3)$$

$$L(\omega, t) = \ln \left( C \left( \omega + \frac{\omega_0}{2}, t \right) - C \left( \omega - \frac{\omega_0}{2}, t \right) \right), \quad (2.4)$$

$$P_{ST}(\omega, t) = e^{\tilde{q}_1(L(\omega+\omega_0,t)+L(\omega-\omega_0,t))+\tilde{q}_0L(\omega,t)}, \quad (2.5)$$

где  $P_{ST}(\omega, t)$  – сглаженная огибающая амплитудного спектра;

$\omega_0 = 2\pi f_0$  – угловая частота основного тона;

$\tilde{q}_0, \tilde{q}_1$  – параметры сглаживания, вычисляемые из спектра оконной функции.

Шум оценивается в виде набора соотношений между периодической и непериодической составляющими сигнала, вычисленных в каждой частотной полосе. Эта информация используется для описания и синтеза невокализованных и частично вокализованных (смешанных) звуков. Оценка шума формируется при помощи линейного предсказания предыдущего и последующего периодов основного тона.

Кратковременный спектр сигнала  $S(\omega, t)$  оценивается следующим образом: период основного тона, выделенный из сигнала, умножается на оконную функцию и дополняется нулями до длины, равной целой степени двойки, затем от полученной последовательности отсчетов вычисляется БПФ. При оценке значений СПМ  $P_T(\omega, t)$  и огибающей  $P_{ST}(\omega, t)$  используется 1,5 периода основного тона, причем предполагается, что параметры сигнала на этом интервале не изменяются. Фактическая продолжительность окна анализа в секундах зависит от частоты основного тона. Влияние изменения частоты основного тона на интервале анализа не учитывается.

Синтез выходного речевого сигнала выполняется на основе метода сложения с перекрытием периодов основного тона, синтезированных в виде минимально-фазовых последовательностей. На рисунке 2.2 показана схема синтеза TANDEM-STRAIGHT.

Огибающая амплитудного спектра используется для формирования минимально-фазовой последовательности, которая моделирует период основного тона, начинающийся размыканием голосовых связок. Для получения минимально-фазовой последовательности необходимо, чтобы логарифм модуля ее спектра и фаза составляли пару преобразования Гильберта.

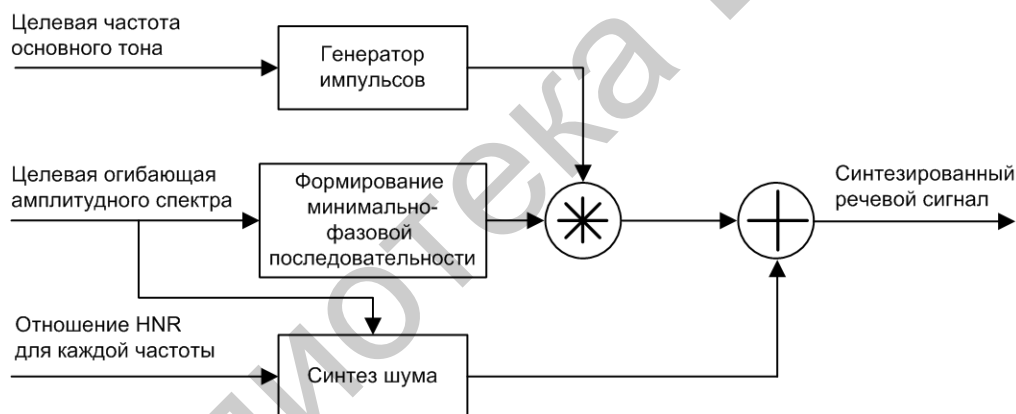


Рисунок 2.2 – Схема синтеза речевого сигнала, используемая в системе параметрической обработки речи TANDEM-STRAIGHT

Процедура получения минимально-фазовой последовательности  $h_{ST}(n)$  состоит из следующих шагов:

1. Вычисление кепстра, используя обратное дискретное преобразование Фурье (ОБПФ) следующим образом:  $C_{ST}(c) = \text{ОБПФ}(\ln(|P_{ST}(\omega, t)|))$ .

2. Обнуление кепстральных коэффициентов для  $c > \frac{N}{2}$  (где  $N$  – порядок преобразования Фурье) и умножение на 2 всех, кроме первого:



$$\bar{c}_{ST}(c) = \begin{cases} C_{ST}(c), & c = 1, \\ 2C_{ST}(c), & 1 < c \leq \frac{N}{2}, \\ 0, & c > \frac{N}{2}. \end{cases}$$

3. Вычисление Фурье-образа последовательности с минимально-фазовой характеристикой  $\bar{H}_{ST}(k) = e^{\text{БПФ}(\bar{c}_{ST}(c))}$ .

4. Получение искомой последовательности из Фурье-образа  $h_{ST}(n) = \text{ОБПФ}(\bar{H}_{ST}(k))$ .

В соответствии с целевым значением частоты основного тона формируется последовательность импульсов, соответствующих моментам смыкания голосовых связок. Для синтеза вокализованной части сигнала вычисляется свертка импульса с текущей последовательностью  $h_{ST}(n, t)$  и результат свертки складывается с перекрытием. Невокализованная часть сигнала синтезируется в виде шума. Выходной сигнал представляет собой смесь вокализованной и невокализованной составляющих.

Модель анализа и синтеза, используемая в системе TANDEM-STRAIGHT, является основной альтернативой синусоидальному моделированию, поскольку обеспечивает относительно высокое субъективное качество реконструированного сигнала. Преимуществом модели также является низкая вычислительная сложность. Оценка параметров модели выполняется на коротких интервалах, что ослабляет эффект усреднения параметров по времени. Однако отсутствие возможности разделения гармонических составляющих и вычисления их параметров приводит к значительной погрешности при оценке огибающей амплитудного спектра. Ограничением модели также является слишком грубый способ формирования сигнала возбуждения. Кроме того, сложение с перекрытием синтезированных периодов основного тона приводит к появлению искажений спектра синтезируемого сигнала.

Синусоидальное моделирование является теоретически более перспективным, поскольку позволяет выполнять анализ отдельных составляющих сигнала и оценивать огибающую амплитудного спектра более точно. Использование синусоидального моделирования также предоставляет более широкие возможности для синтеза сигнала возбуждения. Однако использование синусоидальной модели для параметрического анализа и синтеза речевых сигналов является более сложным, поскольку требует реализации процедур оценки параметров каждой гармоники в условиях изменяющейся частоты основного тона. Для выделения синусоидальных параметров в общем случае требуется больший интервал наблюдения, что ухудшает точность анализа из-за нестационарности сигнала. В этом контексте высокую значимость приобретают методы, позволяющие оценивать мгновенные параметры частотных составляющих с изменяющейся частотой.

## **2.2 Система гибридного моделирования речевого сигнала с изменяющимися синусоидальными параметрами**

В данном подразделе описывается разработанная система параметрического моделирования речевого сигнала на основе синусоидальной модели с нестационарными параметрами, обеспечивающая высокое качество реконструкции выходного сигнала. Субъективно высокое качество реконструкции обеспечивается за счет выполнения анализа, синтеза и обработки квазипериодической части сигнала в измененном масштабе времени, что позволяет точно оценивать параметры модели и синтезировать выходной сигнал без эффекта наложения спектров.

### **2.2.1 Процедура анализа**

Схема процедуры анализа приведена на рисунке 2.3. Анализ состоит из следующих основных шагов:

1. Оценка мгновенной частоты основного тона при помощи устойчивого к ошибкам алгоритма слежения за мгновенной частотой основного тона IRAPT, изложенного выше.

2. Временное масштабирование сигнала для обеспечения стационарности частоты основного тона.

3. Оценка мгновенных гармонических параметров речевого сигнала с использованием ДПФ-модулированного банка фильтров; каждая гармоника основного тона вокализованной речи попадает в отдельный канал банка фильтров, где преобразуется в аналитический комплексный сигнал, из которого выделяется мгновенная амплитуда, фаза и частота согласно выражениям (1.7)–(1.9).

4. На основе анализа полученных значений мгновенной частоты субполосные сигналы классифицируются как периодические и аperiodические.

5. Гармоники, принадлежащие периодическим областям спектра, синтезируются согласно выражению и вычитаются из исходного сигнала.

6. Остаток переводится в частотную область при помощи кратковременного преобразования Фурье.

7. Параметры синтезированных гармоник и СПМ остатка объединяются в одну общую спектральную огибающую.

8. Амплитудные спектры смежных фреймов масштабированного сигнала анализируются для определения способа возбуждения всего анализируемого фрагмента сигнала.

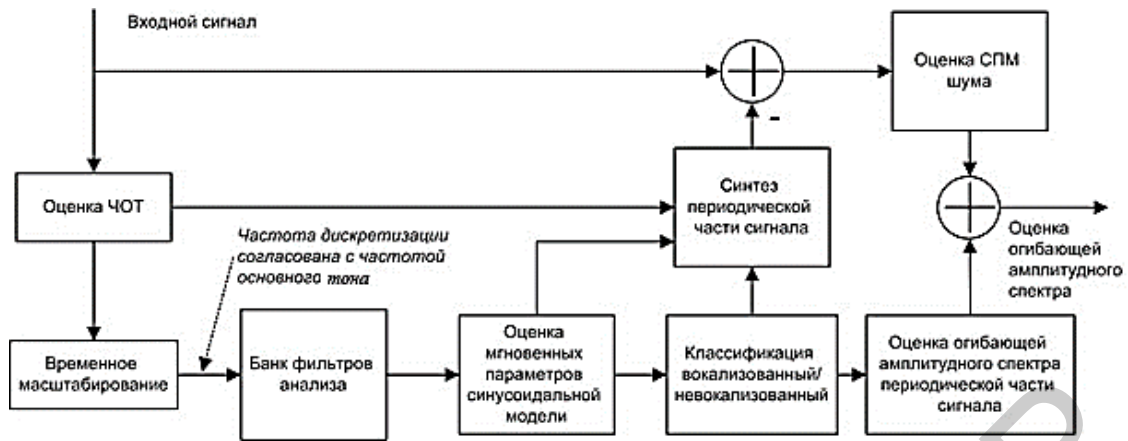


Рисунок 2.3 – Схема оценки параметров синусоидальной модели для параметрического моделирования речевого сигнала

Каждая огибающая амплитудного спектра представляется в виде вектора логарифмических значений энергии, расположенных равномерно в шкале мелов. После временного масштабирования основной тон становится постоянным и каждая гармоника частоты основного тона имеет фиксированное положение в частотном диапазоне.

Каждому входному отсчету речевого сигнала  $s(n)$  ставится в соответствие фаза периода основного тона  $\varphi(n)$ . Новые моменты времени  $m$ , в которые необходимо перерасчитать входной сигнал, показаны на рисунке 2.4.

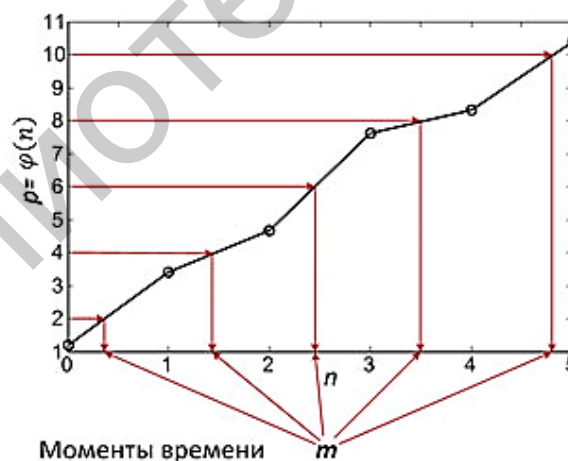


Рисунок 2.4 – Расчет новых моментов времени для масштабирования сигнала

Вычисление значений сигнала в произвольный момент времени  $t$  (рисунок 2.5) выполняется по теореме Котельникова, согласно которой

$$s(t) = \sum_{n=-\infty}^{\infty} s(nT) \operatorname{sinc}\left(\frac{t - nT}{T}\right), \quad (2.6)$$

где  $\operatorname{sinc}(x) = \frac{\sin \pi x}{\pi x}$ ;

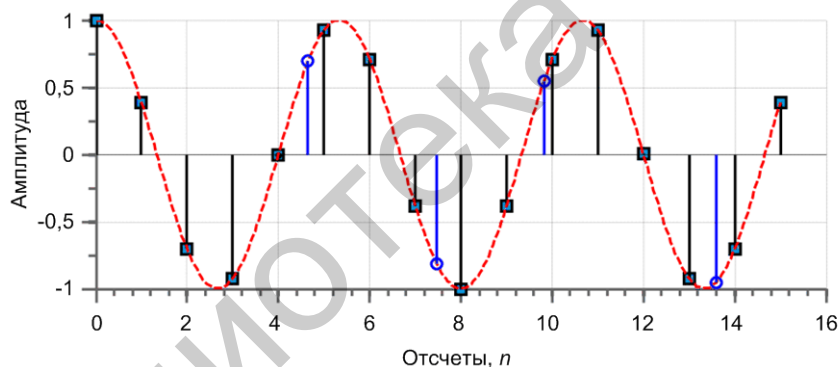
$T$  – интервал дискретизации;

$x(nT)$  – значения сигнала в дискретные моменты времени  $nT$ .

Для простоты в дальнейшем считается, что  $T = 1$  и вместо  $s(nT)$  используется запись  $s(n)$ . Выражение (2.6) нельзя использовать на практике из-за суммирования в бесконечных пределах. Поэтому выбирается конечное число точек сигнала  $N_{pt}$ , предшествующих моменту  $t$ , и  $N_{pt}$  последующих точек:

$$s(t) = \sum_{n=-N_{pt}}^{N_{pt}} s(n) \operatorname{sinc}(t - n) w(t - n),$$

где  $w(t)$  – оконная функция с центром симметрии в точке 0.



— — — — — непрерывный сигнал;

■ – сигнал дискретизированный с равным временным интервалом;

○ – значения сигнала в моменты времени, не кратные интервалу дискретизации

Рисунок 2.5 – Вычисление значений сигнала в произвольные моменты времени

Для каждого выходного отсчета временные отметки пересчитываются таким образом, чтобы текущий момент  $t$  всегда попадал в диапазон от 0 до 1. Это позволяет использовать таблицу с заранее рассчитанными значениями функции  $\operatorname{sinc}$ . Использование табличных значений дает возможность существенно сократить вычислительные затраты.

Для разделения гармоник и оценки их параметров используется ДПФ-модулированный банк фильтров анализа, который представляет собой совокупность фильтров  $H_k(z)$ ,  $0 \leq k < M$ , раскладывающих входной сигнал на  $M$  субполосных сигналов. Амплитудно-частотные характеристики фильтров формируются путем сдвига в частотной области АЧХ фильтра-прототипа. В результате

полоса частот от нуля до частоты дискретизации перекрывается гребенкой из  $M$  фильтров, как показано на рисунке 2.6.

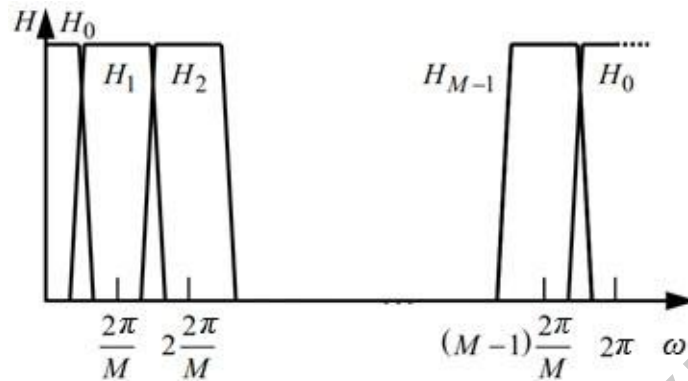


Рисунок 2.6 – АЧХ каналов ДПФ-модулированного банка фильтров

Число каналов в банке выбирается равным  $N_{f_0}$ , поскольку в этом случае центральная частота каждого канала совпадает с положением соответствующей гармоники масштабированного сигнала  $s(m)$ . Импульсные характеристики фильтров анализа определяются выражением

$$h_k(n) = h(n)W_M^{-kn}, \quad k = 0, 1, \dots, M - 1,$$

где  $h(n)$  – фильтр-прототип нижних частот с нормированной частотой среза  $\pi/M$ ;

$$W_M = e^{-j2\pi/M}.$$

Передаточная функция каждого фильтра анализа выражается через передаточную функцию фильтра-прототипа следующим образом:

$$H_k(z) = \sum_{n=0}^{N-1} h_k(n)z^{-n} = \sum_{n=0}^{N-1} h(n)(zW_M^{-k})^{-n} = H(zW_M^{-k}). \quad (2.7)$$

Фильтрация сигнала с использованием импульсных характеристик (2.7) неэффективна, поскольку приводит к большой вычислительной нагрузке. Более эффективный способ фильтрации можно получить, если воспользоваться полифазным представлением фильтра-прототипа, в котором импульсная характеристика  $h(n)$  перегруппирована в  $M$  подпоследовательностей  $e_l(n)$ :

$$H(z) = \sum_{l=0}^{M-1} z^{-l} E_l(z^M), 0 \leq l < M, \quad (2.8)$$

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

Далее объединим выражения (2.7) и (2.8). С учетом тождества  $W_M^{kM} = 1$  получим

$$H_k(z) = \sum_{n=0}^{M-1} z^{-l} W_M^{-lk} E(z^M W_M^{kM}) = \sum_{l=0}^{M-1} z^{-l} W_M^{-kl} E_l(z^M). \quad (2.9)$$

Уравнение (2.9) можно записать в виде произведения вектора строки на вектор-столбец:

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

Все  $M$  уравнений (для каждого  $H_k$ ) в матричной форме можно записать следующим образом:

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

Данное выражение является основой для построения эффективной полифазной структуры ДПФ-модулированного банка фильтров, поскольку матрица  $M \times M$  в правой части является матрицей дискретного преобразования Фурье (ДПФ). Структура банка, использованная в настоящем пособии, приведена на рисунке 2.7.

При реализации ДПФ-модулированного банка фильтров порядок фильтра-прототипа выбирался равным  $N = kM + 1$ , где  $k$  – некоторое целое число.

Выходы банка фильтров  $s_k(m)$  являются узкополосными аналитическими сигналами, которые интерпретируются как отдельные синусоидальные компоненты модели. Параметры синусоидальных компонент вычисляются с интервалом 5 мс при помощи следующих выражений:

$$A_k(m) = \sqrt{\operatorname{Re}^2(s_k(m + \Delta)) + \operatorname{Im}^2(s_k(m + \Delta))},$$

$$\varphi_k(m) = \arg s_k(m + \Delta),$$

$$f_k(m) = \frac{\arg s_k(m + \Delta + 1) - \arg s_k(m + \Delta)}{2\pi},$$

где  $\Delta = \frac{N-1}{2}$  – групповая задержка банка фильтров.

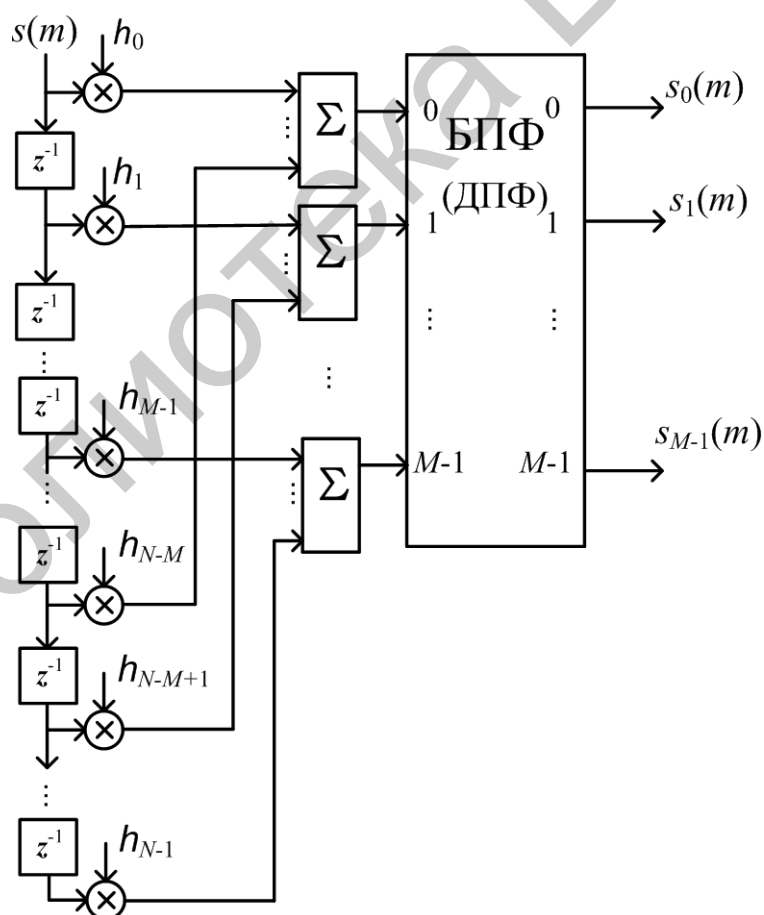


Рисунок 2.7 – Схема эффективной реализации ДПФ-модулированного банка фильтров анализа

На основе полученных параметров выполняется определение признаков вокализованности отдельных гармоник и классификация типа возбуждения, которое может быть полностью вокализованным, полностью невокализованным либо смешанным.

В результате выполнения всей процедуры анализа формируется параметрическое описание сигнала в виде мгновенных гармонических параметров, амплитудной огибающей, шумовой составляющей, классификации регионов спектра по признаку периодичности и типа возбуждения (рисунок 2.8).

### 2.2.2 Процедура синтеза сигнала из параметров модели

Мгновенные фазовые значения изменяются таким образом, чтобы сохранить их непрерывность и синхронизацию фазы между гармониками:

$$\bar{\varphi}_k(m) = \bar{\varphi}_k(m - 1) + \frac{\bar{F}_0(m)}{F_s} 2\pi\Delta t + \varphi_k(m) - kF_0(m),$$

где  $F_s$  – частота дискретизации;

$\Delta t$  – шаг (в отсчетах) между двумя последовательными оценками параметров модели.

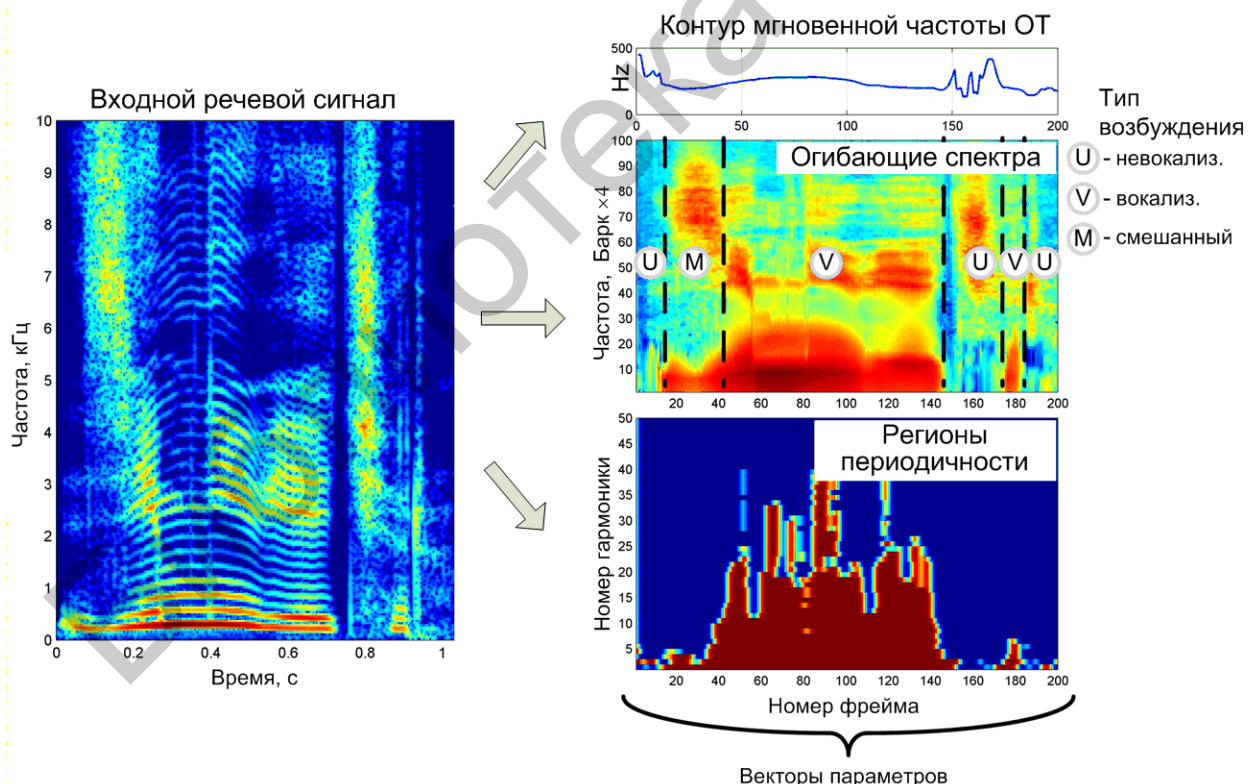


Рисунок 2.8 – Параметризация речевого сигнала гибридной моделью с изменяющимися параметрами



Восстановление сигнала из параметрического представления выполняется путем синтезирования децимированных субполосных сигналов  $\hat{s}_k(m)$  из измененных параметров для каждой гармоники. Затем сигналы подаются на вход банка фильтров синтеза, схема которого показана на рисунках 2.9 и 2.10. Через  $\uparrow S$  обозначен экспандер, добавляющий  $S - 1$  нулевых значений между последовательными отсчетами сигнала  $\hat{s}_k(m)$ . Предполагается, что число  $S$  является делителем  $M$ . Через  $F_k(z)$  на схеме обозначены фильтры синтеза, импульсные характеристики которых определяются следующим образом:

$$f_k(n) = h(n)W_M^{-kn}, \quad k = 0, 1 \dots M - 1, \quad n = 0, 1 \dots N - 1. \quad (2.10)$$

Прямая реализация данной схемы требует значительных вычислительных затрат. Ее можно упростить, если учесть способ получения импульсных характеристик фильтров (2.10) и то, что большинство поступающих в фильтры отсчетов являются нулевыми. Полифазное представление фильтра-прототипа имеет вид

$$H(z) = \sum_{l=0}^{M-1} z^{-(M-1-l)} R_l(z^M), \quad 0 \leq l \leq M, \quad (2.11)$$

где  $R_l(z) = E_{M-1-l}(z)$ .

Объединяя (2.10) и (2.11), получим

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

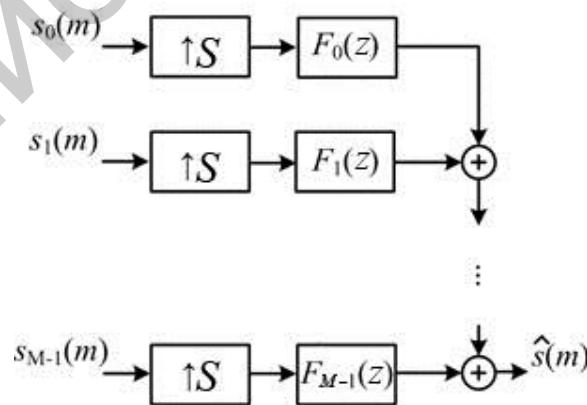


Рисунок 2.9 – Банк фильтров синтеза

Выражение для банка фильтров синтеза можно представить в матричной форме:

$$\hat{S}(z) = [F_0(z) \quad F_1(z) \quad F_2(z) \quad \dots \quad F_{M-1}(z)] \begin{bmatrix} S_0(z) \\ S_1(z) \\ S_2(z) \\ \vdots \\ S_{M-1}(z) \end{bmatrix}.$$

Его же можно записать в следующем виде:

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

$$\times \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & W_M^{-1} & W_M^{-2} & \dots & W_M^{-(M-1)} \\ 1 & W_M^{-2} & W_M^{-4} & \dots & W_M^{-2(M-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_M^{-(M-1)} & W_M^{-2(M-1)} & \dots & W_M^{-(M-1)^2} \end{bmatrix} \times \begin{bmatrix} S_0(z) \\ S_1(z) \\ S_2(z) \\ \vdots \\ S_{M-1}(z) \end{bmatrix}.$$

В этом выражении присутствует матрица размером  $M \times M$  обратного дискретного преобразования Фурье (ОДПФ). В соответствии с этим строится эффективная полифазная структура ДПФ-модулированного банка фильтров синтеза, показанная на рисунке 2.10.

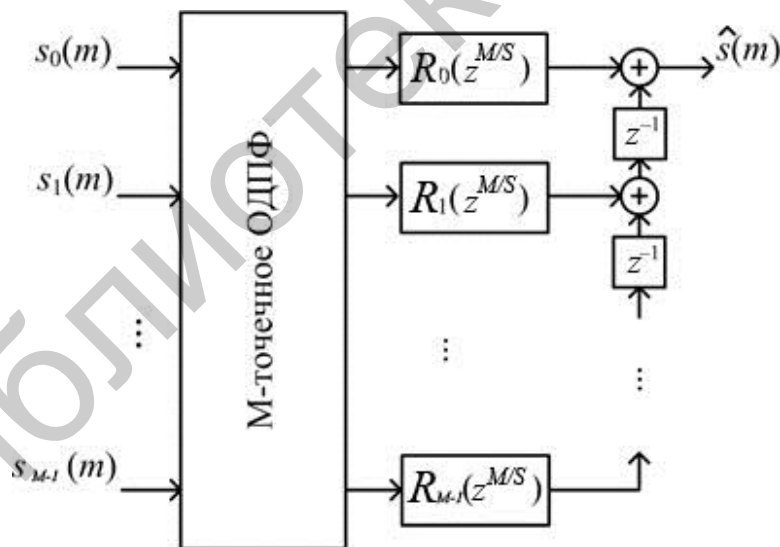


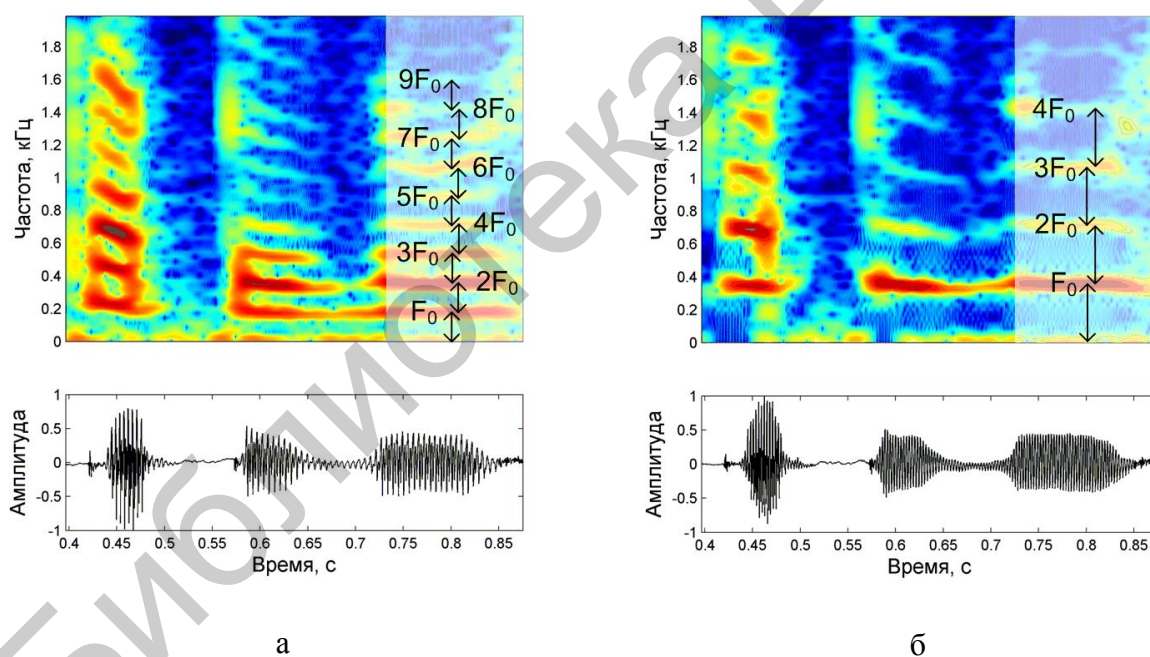
Рисунок 2.10 – Эффективная реализация банка фильтров синтеза

После реконструкции сигнала  $\hat{s}(m)$  выполняется его обратное временное масштабирование с учетом целевого контура основного тона, в результате чего формируется выходной сигнал  $\hat{s}(n)$  с постоянным шагом дискретизации. В аппаратной реализации для синтеза выходного сигнала вместо банка фильтров возможно использование синусоидальных генераторов.

Предложенная схема анализа/синтеза выполняет параметрическое описание нестационарного речевого сигнала в виде суммы компонент с узким спектром, что обеспечивает, с одной стороны, высокую эффективность реализации вычислений за счет прореживания параметров, а с другой – высокое качество реконструкции за счет ослабления эффекта наложения спектров.

### 2.3 Реализация системы изменения частоты основного тона в реальном масштабе времени

Модификация частоты основного тона используется во многих приложениях, таких, как изменение (коррекция) акцента, повышение привлекательности голоса, коррекция певческого голоса, а также при создании многих звуковых эффектов. Синтез голоса с измененным основным тоном является необходимой алгоритмической процедурой для создания синтезаторов речи по тексту, а также синтеза певческого голоса. Изменение частоты основного тона является основным элементом при моделировании экспрессии и эмоций. На рисунке 2.11 показан пример изменения частоты основного тона речевого сигнала.



а – исходный речевой сигнал; б – обработанный сигнал

Рисунок 2.11 – Изменение основного тона речевого сигнала

Эффект изменения частоты основного тона звукового сигнала может достигаться различными методами. Самым простым из них является изменение скорости воспроизведения, что вызывает смещение частоты всех составляющих сигнала. Однако, во-первых, это приводит к изменению длительности сигнала и

поэтому не может быть использовано в приложениях, работающих в реальном масштабе времени, а во-вторых – сильно искажает тембр голоса.

Одной из наиболее популярных альтернатив является фазовый вокодер и разнообразные методы на его основе, выполняющие смещение компонент сигнала в частотной области при помощи прямого и обратного преобразования Фурье. Применение фазового кодера позволяет изменять высоту звучания сигнала без изменения длительности и сохранять тембр путем коррекции спектральной огибающей. Данный подход может быть использован для любых звуковых сигналов. Вокодер может быть реализован в режиме реального времени, поскольку выполняет обработку входного сигнала последовательно фрейм за фреймом. При применении его к речи и певческому голосу метод имеет существенные ограничения: при обработке не используется какая-либо модель голосообразования, что при значительном изменении частоты основного тона приводит к неестественному звучанию; выделяемые частотные составляющие сигнала не соответствуют гармоникам основного тона, что приводит к потере натуральности и звонкости голоса; сигнал не разделяется на вокализованный и невокализованный, что приводит к неестественному звучанию некоторых звуков (в частности, невокализованных шипящих «с», «ш», а также смешанных, частично вокализованных «з» и «ж»).

Для повышения натуральности звучания и расширения доступного диапазона изменения основного тона необходимо использовать более сложные решения, основанные на гибридной (детерминистской/стохастической) модели речевого сигнала. Гибридная обработка звуковых сигналов предложена в работе, где используется три отдельные составляющие: периодическая, шумовая и транзистная. Поскольку при изменении основного тона голоса обрабатывается только вокализованная (детерминистская) часть сигнала, в данном случае нет необходимости различать между собой шумовые и транзистные составляющие – вместе их можно отнести к невокализованной (стохастической) части сигнала.

Рассмотренный выше способ параметризации, основанный на использовании мгновенных синусоидальных параметров, обеспечивающий высококачественный синтез сигнала с измененной просодикой, позволяет предложить решение, которое приводится в настоящем подразделе. Решение основано на гибридном представлении сигнала в форме синусоиды плюс остаток. Частотно-временной анализ выполняется с учетом контура частоты основного тона, что позволяет выделять параметры отдельных гармоник и разделять сигнал на вокализованную и невокализованную составляющие в частотной области. Моделирование вокализованного сигнала выполняется при помощи нестационарных параметров. В экспериментальной части работы приводятся результаты субъективной оценки качества обработанной речи. С целью оценки применимости метода в существующих каналах связи для экспериментов использовались как широкополосные, так и узкополосные речевые сигналы с применением кодеков G.711 и GSM.

Алгоритм обработки сигнала можно кратко описать в виде последовательности действий:

- определение частоты основного тона  $F_0(n)$ ;
- преобразование речевого сигнала в параметрический вид, оценка мгновенных гармонических параметров  $A_k(n)$ ,  $f_k(n)$  и  $\varphi_k(n)$ ,  $k = 1, 2, \dots, K$ ;
- оценка вокализации каждой тройки гармонических параметров и отбор только тех, которые относятся к вокализованным областям спектра;
- синтез исходной вокализованной компоненты сигнала и ее вычитание из исходного речевого сигнала для получения шумовой составляющей  $r(n)$ ;
- синтез вокализованной компоненты с измененным основным тоном и ее сложение с шумовой составляющей  $r(n)$ .

Оценка параметров модели выполняется с учетом контура мгновенной частоты основного тона. Входной речевой сигнал масштабируется во времени для того, чтобы обеспечить его стационарность, а затем выполняется узкополосная фильтрация, разделяющая гармоники основного тона. В результате фильтрации формируются аналитические (комплексные) сигналы, которые описываются при помощи параметров синусоидальной модели. На основе анализа смежных значений мгновенной частоты определяется степень вокализации каждой гармоники. Затем выполняется синтез квазипериодического сигнала по полученным параметрам, который вычитается из исходного речевого сигнала. В результате выполняется разделение речевого сигнала на вокализованную и невокализованную части. Одновременно синтезируется квазипериодический сигнал с целевой частотой основного тона, который складывается с полученной невокализованной частью.

Целевой контур основного тона формируется согласно требованиям конкретного приложения. Для экспериментов использовались заданные профили изменения исходного основного тона, приведенные в соответствующем разделе.

Полученные параметры гармонической модели  $A_k(n)$ ,  $\varphi_k(n)$  и  $f_0(n)$  изменяются в соответствии с новым значением мгновенной частоты основного тона  $\bar{f}_0(n)$ . В каждый момент времени амплитуды гармоник входного сигнала  $A_k$  определены на сетке частот  $kf_0$ . Поскольку основной тон изменяется, но сохраняется исходная спектральная огибающая, то необходимо перерасчитать амплитуды на новой частотной сетке  $k\bar{f}_0$ , как показано на рисунке 2.12. Перерасчет осуществляется при помощи линейной интерполяции.

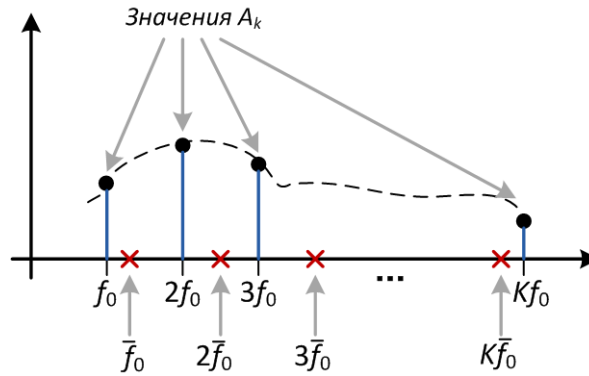


Рисунок 2.12 – Вычисление амплитудных значений гармоник новой частоты основного тона

Частота основного тона определяет высоту звучания голоса. Контур частоты основного тона (его изменение в зависимости от времени на большом протяжении) определяет интонацию речи.

Чтобы сохранить исходный тембр диктора при изменении основного тона, должна сохраняться спектральная огибающая, которая формируется из мгновенных амплитуд гармоник основного тона. Спектральную огибающую можно рассматривать как функцию от номера отсчета и частоты  $E(n, f)$ , которая принимает значения мгновенных амплитуд гармоник основного тона в соответствующих точках  $E(n, f_k(n)) = A_k(n)$ . Для произвольных  $n$  и  $f$  функция вычисляется путем линейной интерполяции ближайших к ним амплитудных значений.

Синтез голоса с модифицированным контуром частоты основного тона может быть выполнен по следующей формуле:

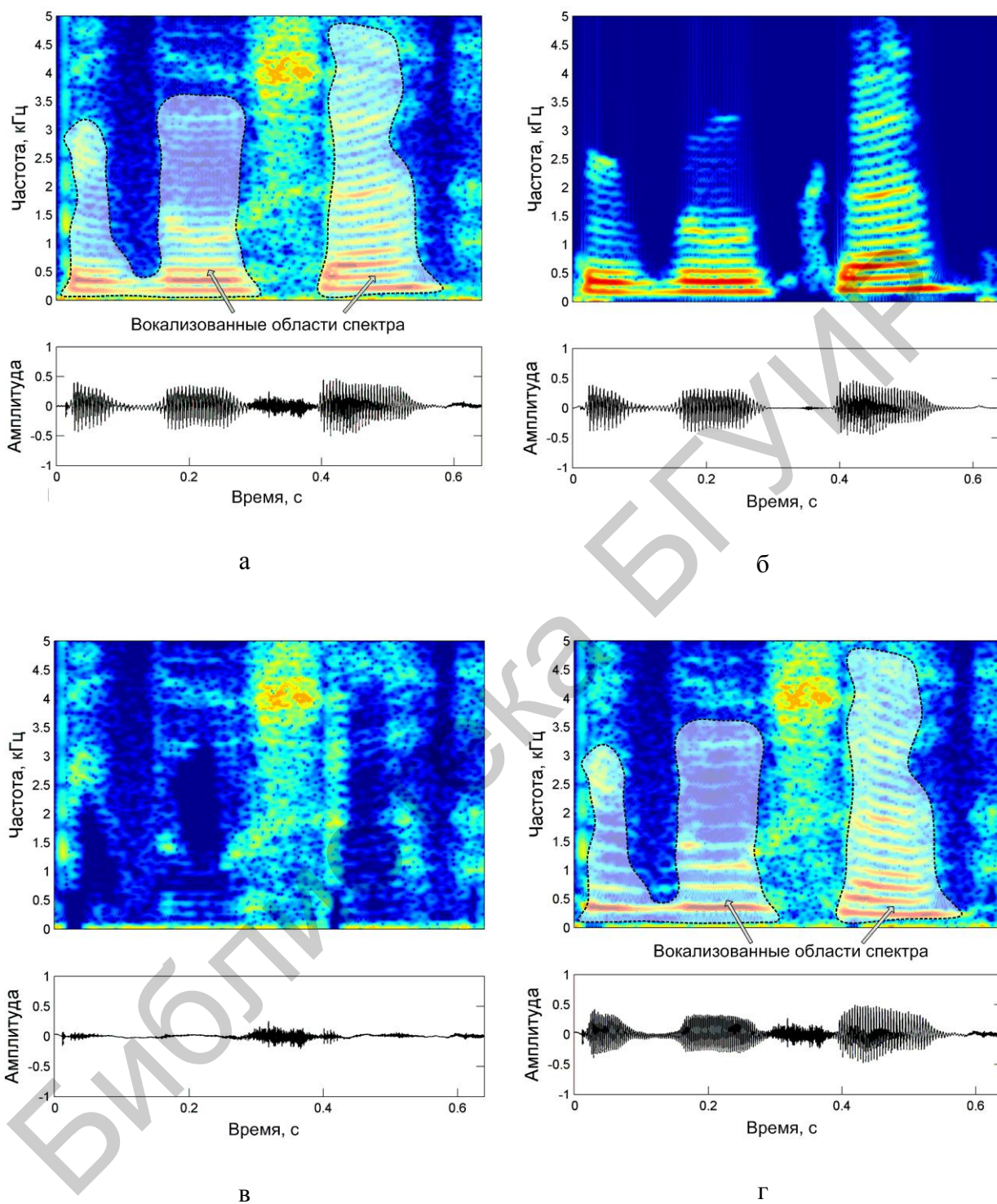
$$s(n) = \sum_{k=1}^K E(n, \bar{f}_0(n)k) \cos \bar{\varphi}_k(n) + r(n),$$

где фазы гармонических компонент  $\bar{\varphi}_k(n)$  рассчитываются в соответствии с новым контуром частоты основного тона  $\bar{f}_0(n)$  следующим образом:

$$\bar{\varphi}_k(n) = \sum_{i=0}^n \frac{2\pi \bar{f}_0(i)}{F_s} + \bar{\varphi}_k^\Delta(n).$$

Дополнительный фазовый параметр  $\bar{\varphi}_k^\Delta(n)$  используется для сохранения относительных фаз гармоник по отношению к фазе частоты основного тона. Данный параметр вычисляется как

$$\bar{\varphi}_k^\Delta(n) = \varphi_k(n) - k\varphi_0(n).$$



а – исходный речевой сигнал; б – выделенная вокализованная часть;  
 в – сигнал-остаток  $r(n)$ ; г – обработанный выходной сигнал

Рисунок 2.13 – Разделение сигнала на периодическую и стохастическую составляющие

Основной тон присутствует только в вокализованных сегментах речи, т. е. только тогда, когда задействованы голосовые связки диктора. Такие звуки, как «а», «о», «ж», являются вокализованными, в то время как звуки «с», «ш», «щ» являются невокализованными. Как было показано выше, в спектре речевого сигнала вокализованность проявляется в виде спектральных компонент кратной частоты. Чтобы сохранить исходное качество звучания невокализованных звуков, алгоритм изменения тона должен автоматически выделять области вокализованности и выполнять обработку только в этих областях. Шумовая часть  $r(n)$  выделяется из исходного сигнала вычитанием выделенных вокализованных звуков, как показано на рисунке 2.13.

Оценка основного тона выполняется при помощи алгоритма IRAPT. Особенностью алгоритма является возможность определения мгновенной частоты, которая достигается использованием специальной функции оценки периодичности аналогичной автокорреляционной функции, вычисляемой из мгновенных гармонических параметров субполосных составляющих сигнала.

Оценка параметров отдельных гармоник выполняется при помощи узкополосной фильтрации. Чтобы повысить частотное разрешение анализа, масштаб времени сигнала изменяется согласованно с полученным контуром мгновенной частоты основного тона. Сигнал дискретизируется таким образом, чтобы на каждый период основного тона приходилось равное количество отсчетов  $N_{f_0}$ .

#### 2.4 Результаты экспериментов изменения частоты основного тона речевого сигнала

Для тестирования качества работы системы было разработано четыре эффекта (профиля) изменения частоты основного тона.

Первый эффект использует фиксированное повышение основного тона. Целевая частота основного тона  $\bar{f}_0(n)$  и исходная  $f_0(n)$  связываются соотношением

$$\bar{f}_0(n) = f_0(n)k, \quad k > 1,$$

где  $k$  – коэффициент повышения.

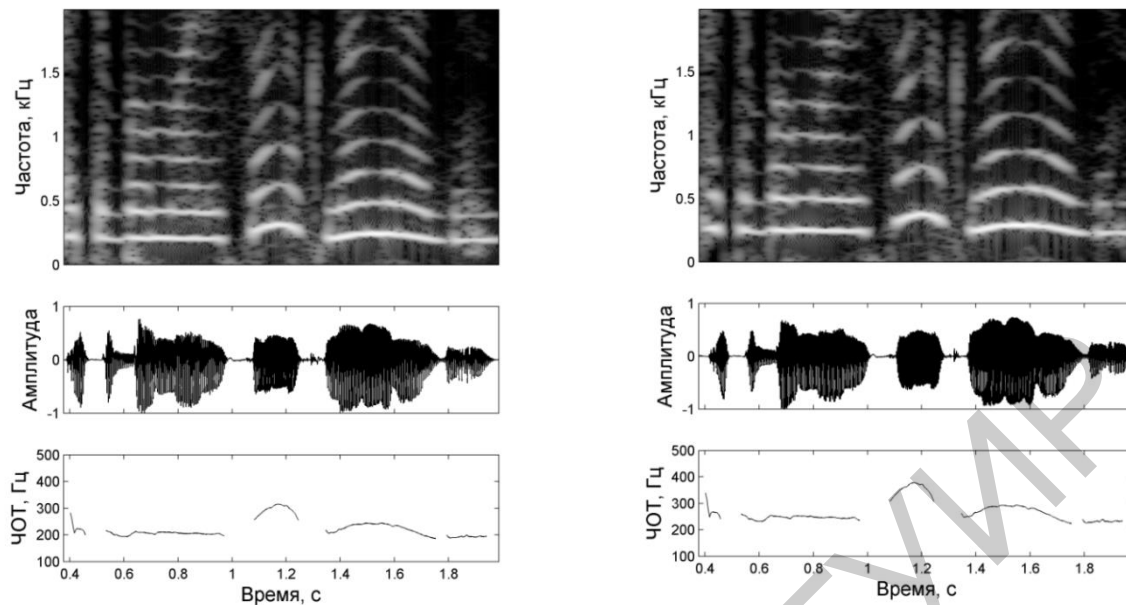
Высокая натуральность синтезированного голоса достигается, когда коэффициент  $k < 2$ . На рисунке 2.14 показан результат повышения основного тона.

Второй эффект заключался в понижении основного тона. Целевая частота основного тона определялась из исходной следующим образом:

$$\bar{f}_0(n) = \frac{f_0(n)}{k}, \quad k > 1,$$

где  $k$  – коэффициент, на который происходит понижение тона.





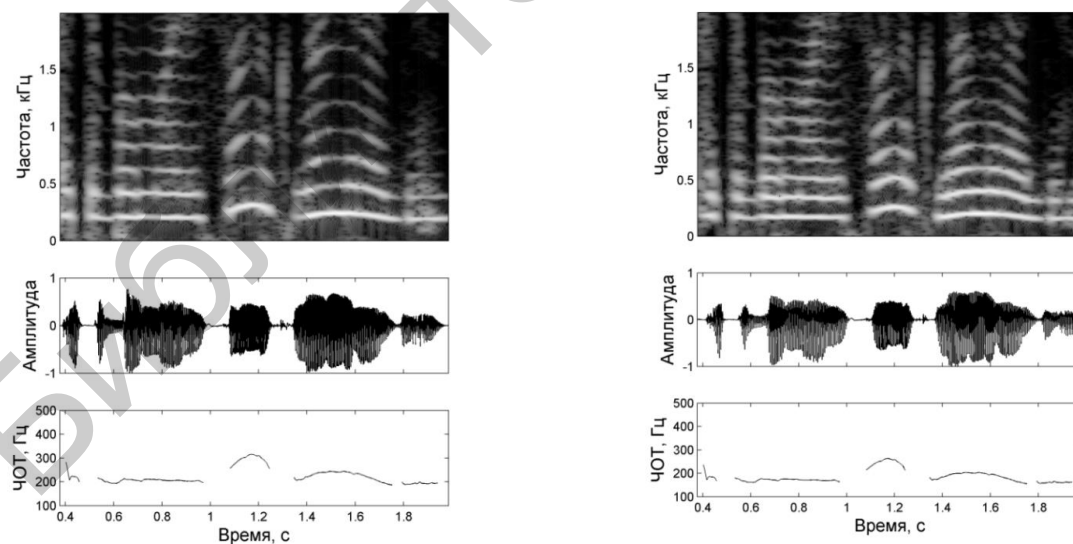
а

б

а – входной сигнал; б – обработанный сигнал

Рисунок 2.14 – Эффект повышения частоты основного тона

Высокая натуральность синтезированного голоса достигается, если коэффициент  $k$  не превышает 2. На рисунке 2.15 показан пример понижения основного тона.



а

б

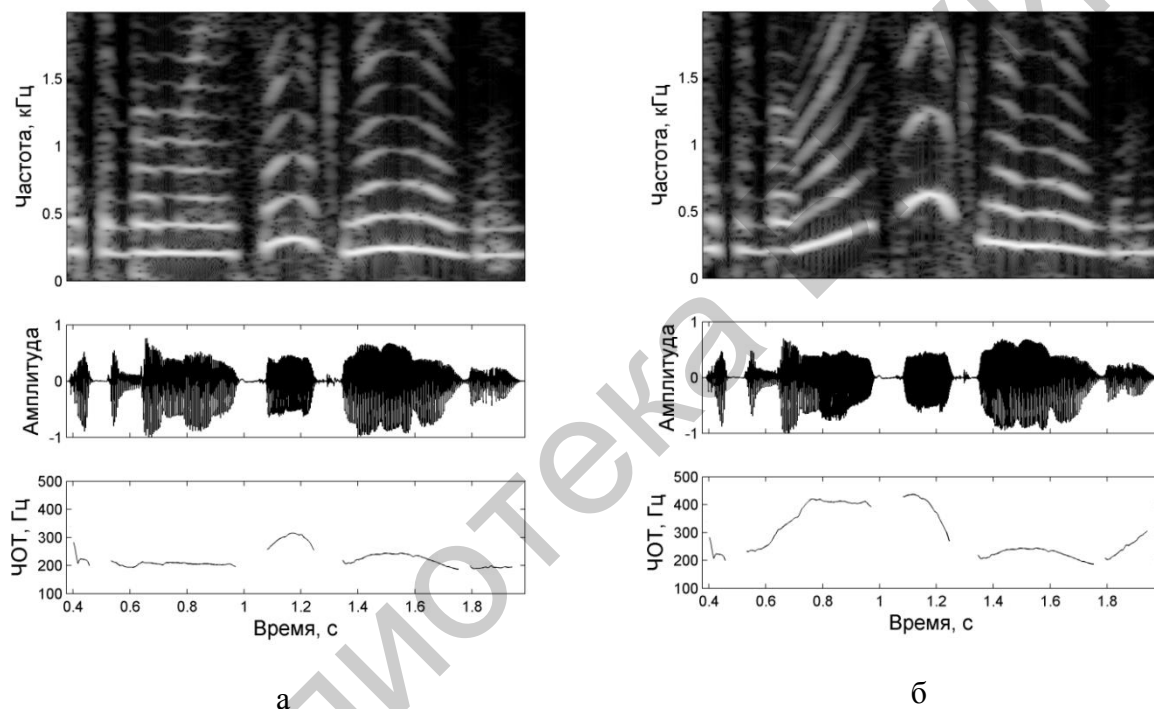
а – входной сигнал; б – обработанный сигнал

Рисунок 2.15 – Эффект понижения частоты основного тона

Третий эффект «*cycle*» выполняет периодическое повышение тона. Профиль, согласно которому происходит повышение, задается функцией  $f_{\text{profile}}(n)$ , определенной на конечном интервале от 0 до  $N - 1$ . При реализации эффекта «*cycle*» целевая частота основного тона формируется следующим образом:

$$\bar{f}_0(n) = f_0(n) \cdot f_{\text{profile}}(n \bmod N).$$

На рисунке 2.16 показан результат применения эффекта «*cycle*» к речевому сигналу.



а – входной сигнал; б – обработанный сигнал

Рисунок 2.16 – Эффект «*cycle*»

Четвертый эффект «*sinus*» позволяет придать голосу дрожание, которое иногда появляется у человека в момент волнения. Для достижения этого эффекта частота основного тона сигнала преобразуется следующим образом:

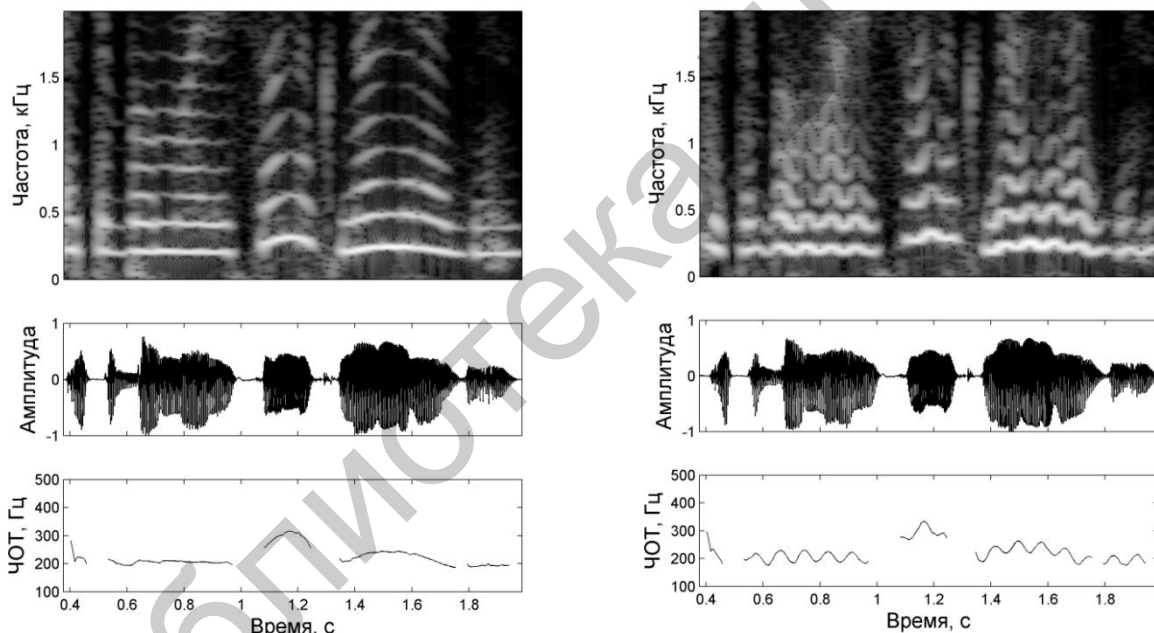
$$\bar{f}_0(n) = f_0(n) + 20 \cos\left(2\pi n \frac{12}{F_s}\right),$$

где  $F_s$  – частота дискретизации сигнала.

На рисунке 2.17 показан результат применения эффекта «*sinus*» к речевому сигналу.

Тестирование системы изменения основного тона речевого сигнала выполнялось на речевой базе, начитанной девятью дикторами (четыре мужских голоса и пять женских). Длительность записи каждого диктора приблизительно составляла 1 мин 10 с. Чтобы оценить качество работы системы в телекоммуникациях, вместе с исходными сигналами обрабатывались сигналы, декодированные кодеками GSM и G.711.

Образцы сигналов, обработанные четырьмя различными эффектами, прослушивались группой экспертов. Задача экспертов заключалась в оценке натуральности речи по 5-балльной шкале (5 – натуральная речь; 4 – речь звучит натурально, заметны отдельные артефакты; 3 – синтезированная речь, имеются артефакты; 2 – ненатуральное звучание, отдельные слова не воспринимаются; 1 – полная потеря разборчивости). В таблице 2.1 приведены средние оценки, полученные в результате эксперимента.



а

б

а – входной сигнал; б – обработанный сигнал

Рисунок 2.17 – Эффект «*sinus*»

На вход системы подавались сигналы трех различных категорий: широкополосная речь, узкополосная речь и узкополосная речь, декодированная после сжатия кодеками GSM и G.711. Для каждой категории сигнала и каждого эффекта изменения тона было синтезировано девять различных образцов по числу дикторов. Каждый из полученных образцов оценивался группой экспертов, а в таблицу заносились только средние значения оценок.

Таблица 2.1 – Оценка натуральности образцов речи, обработанных в системе реального времени

Образец	Звуковой эффект				
	Без обработки	Повышение тона $k = 1,2$	Понижение тона $k = 1,2$	Эффект « <i>cycle</i> »	Эффект « <i>sinus</i> »
Широкополосная речь	5	4,80	4,75	4,30	3,95
Речь после кодека GSM	4,25	4,15	4,20	3,85	3,70
Речь после кодека G.711	4,15	4,05	4,15	3,85	3,75
Речь после кодеков G.711 и GSM	4,10	4,00	4,10	3,80	3,60

Очевидно, что натуральность речи будет тем выше, чем меньше изменений вносится в контур частоты основного тона. Это подтверждается и экспериментальными данными. Эффекты понижения и повышения вносят наименьшие изменения и по этой причине они получили более высокую оценку. Эффект «*cycle*» имеет временные интервалы, на которых динамика изменения контура частоты основного тона остается неизменной. Поэтому натуральность данного эффекта выше, чем у эффекта «*sinus*», который вносит систематическое изменение в натуральную форму контура.

## 2.5 Система коррекции певческого голоса

В последнее время разработчики и пользователи проявляют заметный интерес к решению задачи коррекции певческого голоса для систем различного типа с целью обработки голосов певцов-любителей до уровня профессионального звучания. Известная система синтеза певческого голоса Vocaloid выполняет синтез из слов песни и нотной записи мелодии. Возможности системы расширяет специальный модуль VocaListener, который извлекает характеристики из голоса пользователя и использует их при синтезе, создавая при этом персонифицированное звучание синтезированного вокала. Существуют системы, которые для синтеза используют целевую мелодию и текст, начитанный пользователем. Однако есть потребность в создании системы, которая использует для синтеза целевую мелодию песни и оригинальный певческий голос пользователя, причем пользователь исполняет песню настолько хорошо, насколько это ему доступно, а система корректирует его голос таким образом, чтобы он всегда попадал в мелодию, обеспечивая воспроизведение правильных нот. Целью является сохранение натурального исходного голоса и тонких деталей исполнения пользователя настолько точно, насколько это возможно. Предлагаемое решение основано на системе параметрического моделирования с изменяющимися параметрами. В

настоящее время наиболее популярной системой параметрического моделирования голоса является TANDEM-STRAIGHT, которую удовлетворительно удалось применить к обработке певческого голоса. Однако, благодаря идее выполнения обработки сигнала в измененном масштабе времени, синхронизированном с контуром частоты основного тона, предложенный способ обладает возможностью выделения и моделирования более тонких деталей сигнала, что является преимуществом для обработки певческих голосов, особенно при использовании сложных режимов фонации (глоттализация, хриплый голос, неполное смыкание голосовых связок и т. д.).

### **2.5.1 Формирование целевого контура частоты основного тона для певческого голоса**

Целевой контур частоты основного тона для певческого голоса формируется из нот заданной мелодии с учетом действительного контура частоты основного тона, выделенного из голоса пользователя. В первую очередь, выполняется подбор нужной октавы и транспонирование мелодии в соответствии со средней высотой голоса. Далее выполняется выравнивание мелодии песни по времени по отношению к исполнению пользователя путем ДП, что значительно сокращает число слышимых артефактов в выходном сигнале при переходе от ноты к ноте. При выравнивании используется целевая функция расстояния, учитывающая значение частоты основного тона, а также признаки вокализованности. Далее контур частоты основного тона эвристически сегментируется на ноты, которые притягиваются к мелодии.

Исходная форма контура основного тона сохраняется на границах сегментов для ослабления эффекта «компьютерного акцента». Для сокращения разрывов между исходным и целевым контуром разрешается использование мелодических вариаций. Возможные вариации предопределяются в мелодии путем задания одновременно звучащих разрешенных нот. Поиск целевого контура с учетом вариаций выполняется при помощи ДП. Пример формирования целевого контура частоты основного тона показан на рисунке 2.18.

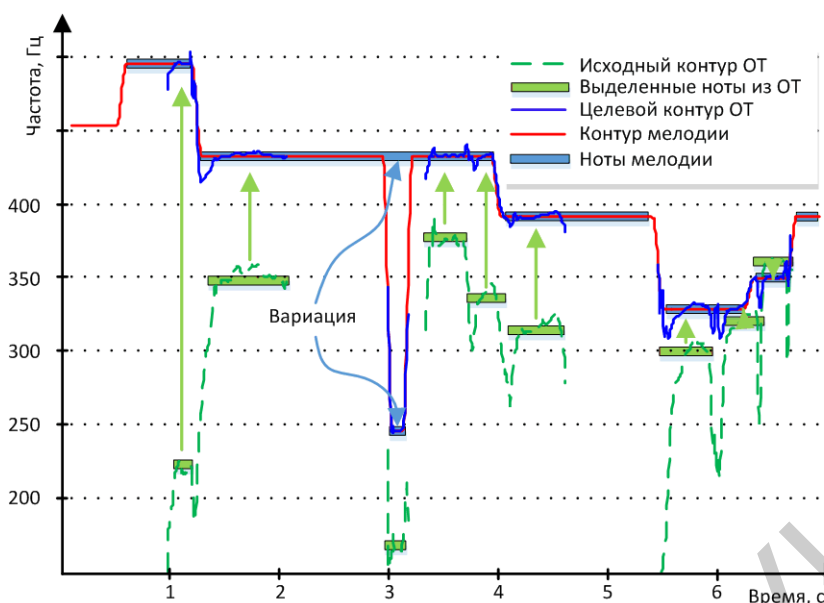


Рисунок 2.18 – Формирование целевого контура частоты основного тона для системы автоматической коррекции певческого голоса

### 2.5.2 Реализация системы автоматической коррекции певческого голоса с удаленной обработкой

Система автоматической коррекции певческого голоса реализована в виде интернет-сервиса. С использованием MATLAB-реализации параметрической обработки голоса создан удаленный сервер, который обрабатывает входные звуковые файлы в соответствии с заданной нотной записью мелодии. Общая схема функционирования системы представлена на рисунке 2.19.

Для записи певческого голоса пользователя и связи с сервером разработано специальное клиентское приложение на мобильном устройстве (смартфоне). Пользовательский сеанс работы с приложением включает два шага.

На первом шаге пользователь поет и выполняет запись своего голоса. Во время пения пользователь слышит музыкальное сопровождение в наушники, слышит свой исходный голос с малой задержкой (менее 20 мс) и имеет возможность читать текст на экране.

На втором шаге выполняется обработка голоса. Контур частоты основного тона выделяется из входного голоса пользователя и затем, используя ноты мелодии, формируется целевой контур частоты основного тона. Выполняется параметризация входного сигнала и модификация параметров в соответствии с целевым контуром. Певческий голос синтезируется с измененной частотой основного тона и результат синтеза микшируется с музыкальным сопровождением. Полученное музыкальное произведение кодируется и передается пользователю для прослушивания.

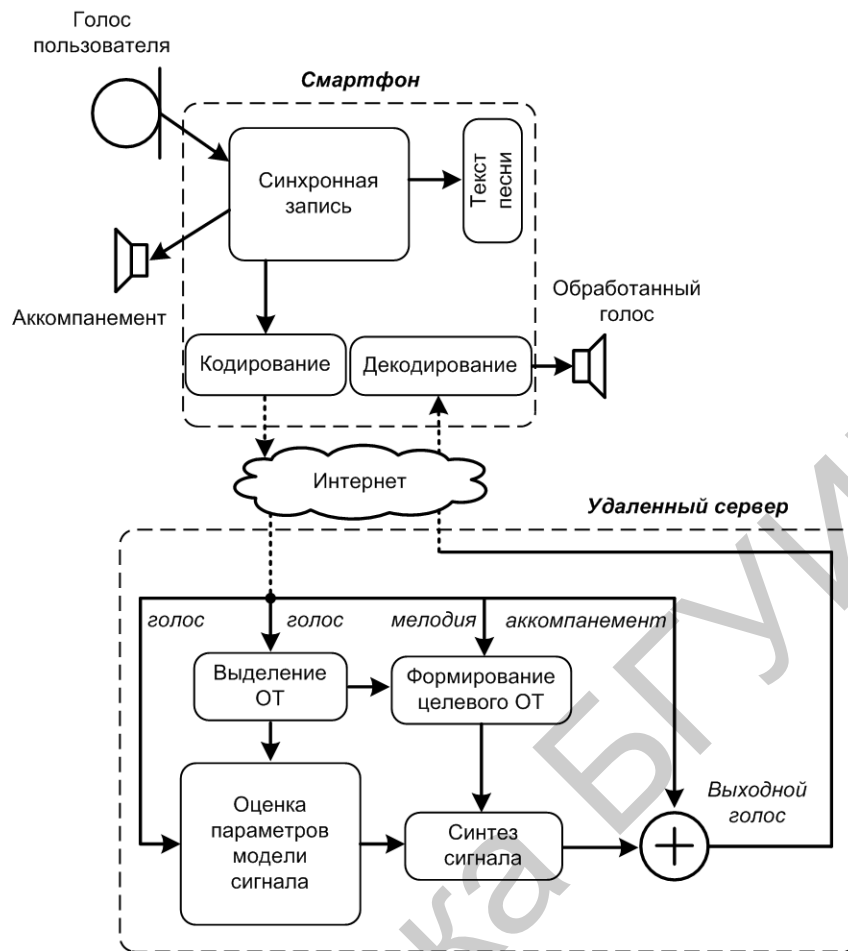


Рисунок 2.19 – Система коррекции певческого голоса с удаленной обработкой

### 3 ПАРАМЕТРИЧЕСКИЙ АНАЛИЗ АУДИОСИГНАЛОВ НА ОСНОВЕ РАЗРЕЖЕННОЙ АППРОКСИМАЦИИ С ПЕРЦЕПТУАЛЬНО-ОПТИМИЗИРОВАННЫМ ПДВП СЛОВАРЕМ ВЕЙВЛЕТ-КОЭФФИЦИЕНТОВ

#### 3.1 Разреженная аппроксимация аудиосигналов. Постановка задачи

Адаптивная частотно-временная аппроксимация аудиосигнала на основе пакетного дискретного вейвлет-преобразования (ПДВП) с перцептуально-оптимизированным частотно-временным планом позволяет исключить избыточность в аудиосигнале, что обусловлено невосприятием определенных частотных компонент человеком, вследствие маскирующего эффекта. Однако использование частотно-временного преобразования на основе алгоритма согласованной подгонки (Matching Pursuit (MP)) со словарем частотно-временных функций позволяет получить разреженную частотно-временную аппроксимацию аудиосигналов, а следовательно, лучшую, чем на основе вейвлет-преобразования, и ПДВП-локализацию нестационарности в сигнале. Это в конечном итоге позволит уменьшить битовый поток в схеме компрессии.

Алгоритм согласованной подгонки обеспечивает разреженную аппроксимацию сигнала на словаре функций:

$$D = \{g_m(n), m = 0..M - 1, n = 0..N - 1, \|g_m\| = 1\}. \quad (3.1)$$

При этом строится квазиоптимальное решение задачи разреженной аппроксимации, т. е. строится модель вида

$$x(n) = \sum_{m=0}^{M-1} \alpha_m \cdot g_{\gamma m}, \quad (3.2)$$

где  $\alpha_m$  – весовой коэффициент для выбранной из словаря частотно-временной функции  $g_{\gamma m}$  на  $m$ -й итерации алгоритма.

Разреженная аппроксимация аудиосигналов на основе алгоритма MP базируется на методе анализа через синтез. Положим, что

$$D = (g_\gamma)_{\gamma \in \Gamma} \quad (3.3)$$

есть полный словарь частотно-временных функций, в котором множество элементов, относительно которых параметризуются частотно-временные функции, например масштабный фактор, частота, время, индексируется  $\gamma \in \Gamma$ , где  $\Gamma$  – случайное множество индексов. Пусть дана некоторая частотно-временная функция



$g_\gamma$ , параметризованная относительно индекса  $\gamma$ , тогда наилучшая возможная аппроксимация  $x$  получается как ортогональная проекция аудиосигнала  $x(n)$  на стянутое подпространство частотно-временной функции  $g_\gamma$  (рисунок 3.1).

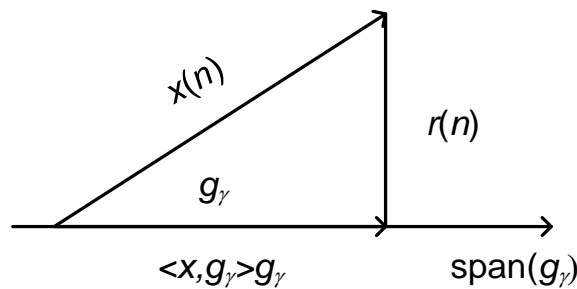


Рисунок 3.1 – Ортогональная проекция  $x(n)$  на  $\text{span}(g_\gamma)$

Таким образом, аудиосигнал может быть декомпозирован как

$$x(n) = \langle x, g_\gamma \rangle g_\gamma + r(n), \quad (3.4)$$

где  $r(n)$  – остаточный сигнал после вычитания проекции  $\langle x, g_\gamma \rangle g_\gamma$ . На основании ортогональности  $r(n)$  и  $g_\gamma$  следует, что

$$\|x(n)\|^2 = |\langle x, g_\gamma \rangle|^2 + \|r(n)\|^2. \quad (3.5)$$

Данная декомпозиция делается для любого и каждого элемента словаря, и лучшее согласование будет найдено путем выбора элемента  $g_\gamma$ , для которого  $\|r(n)\|$  минимальна, или, что эквивалентно, для которого  $|\langle x, g_\gamma \rangle|$  – максимальный. Математически это можно сформулировать следующим образом:

$$\gamma' = \arg \sup_{\gamma \in \Gamma} |\langle x, g_\gamma \rangle|. \quad (3.6)$$

Таким образом, сигнал  $x(t)$  проецируется на избыточный словарь частотно-временных функций со всеми возможными комбинациями масштабов, переходных и модуляционных параметров. Например, на рисунке 3.2 показан пример такой параметрической частотно-временной функции Гаусса. Когда  $x(n)$  реальный и дискретный, как аудиосигнал, тогда используется словарь реальных и дискретных функций. Благодаря избыточности словаря это дает возможность максимально гибко выбрать лучшую частотно-временную функцию для локальной структуры сигнала (локальная оптимизация). Максимальная гибкость модели сигнала позволяет осуществлять компактную аппроксимацию сигнала с

максимально возможной точностью минимально возможным числом частотно-временных функций.

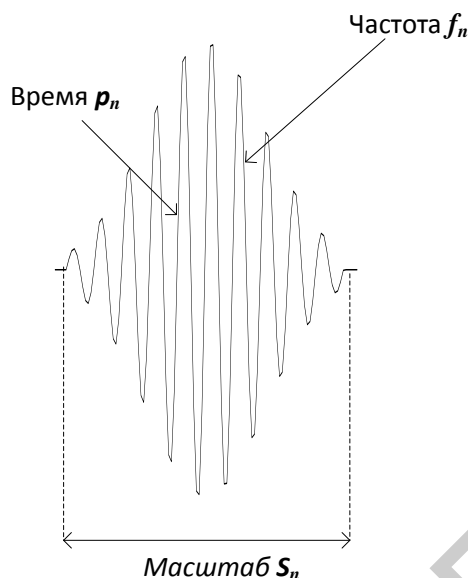


Рисунок 3.2 – Параметрическая частотно-временная функция Гаусса

Главная проблема при практической реализации данного подхода – это построение словаря частотно-временных функций. Наиболее естественным способом формирования словаря является определение частотно-временных функций из самого сигнала. Минимизация вычислительной сложности метода напрямую связана также со структурой и размером словаря. Выбор частотно-временных функций определяется моделью аудиосигнала, а следовательно, и методом анализа и синтеза.

### 3.2 Параметрический анализ аудиосигналов на основе разреженной аппроксимации и полным ПДВП-словарем вейвлет-коэффициентов

Параметрический анализ аудиосигналов базируется на алгоритме МР, который итеративно подбирает аппроксимирующие функции из словаря для описания фрейма входного аудиосигнала. ПДВП как основа для анализа сигналов в алгоритме МР выполняет разложение входного сигнала по вейвлет-функциям, частотные характеристики которых имитируют соответствующие характеристики сигнала. На основе данных вейвлет-функции формируется словарь атомов (ортонормальных векторов декомпозиции), каждый из которых хорошо локализован как по частоте, так и масштабу. Применение полного дерева ПДВП позволяет сформировать избыточный словарь векторов декомпозиции.

Алгоритм МР аппроксимирует анализируемый фрейм сигнала  $x(n)$  атомами  $g_\gamma$ , выбираемыми из словаря  $D$ . 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 частотно-временную локализацию, схожую с дискретной функцией окна, которая увеличивается на  $2^l$  и имеет центр в  $2^l(k + 1/2)$ .

Структура алгоритма разреженной аппроксимации, используя МР со словарем на основе полного дерева ПДВП, показана на рисунке 3.3. Реализация алгоритма МР заключается в итерационном повторении следующих шагов.

Шаг 1. Декомпозиция остаточного сигнала  $r^m(n)$  банком фильтров на основе полного дерева ПДВП  $E_{full}$ , для  $m = 1$  итерации  $r^m(n) = x(n)$ .

Шаг 2. Выбор наиболее значимого вейвлет-коэффициента  $X_\gamma$ , т. е. коэффициента с абсолютным максимальным значением весового коэффициента  $a_\gamma$ .

Шаг 3. Наиболее значимому вейвлет-коэффициенту  $X_\gamma$  ставится в соответствие атом  $g_\gamma$  из словаря  $D$ .

Шаг 4. Формирование результирующего вектора  $g_\gamma^m(n)$  декомпозиции выполняется при помощи обратного ПДВП-преобразования на основании структуры дерева ПДВП  $E_{full}$ .

Шаг 5. Получение результирующего сигнала как умножение результирующего вектора  $g_\gamma^m(n)$  на весовой коэффициент  $a_\gamma^m$ .

Шаг 6. Получение остаточного сигнала  $r^{m+1}(n)$  путем вычитания результирующего сигнала  $(a_\gamma^m \cdot g_\gamma^m(n))$  из остаточного сигнала  $r^m(n)$ .

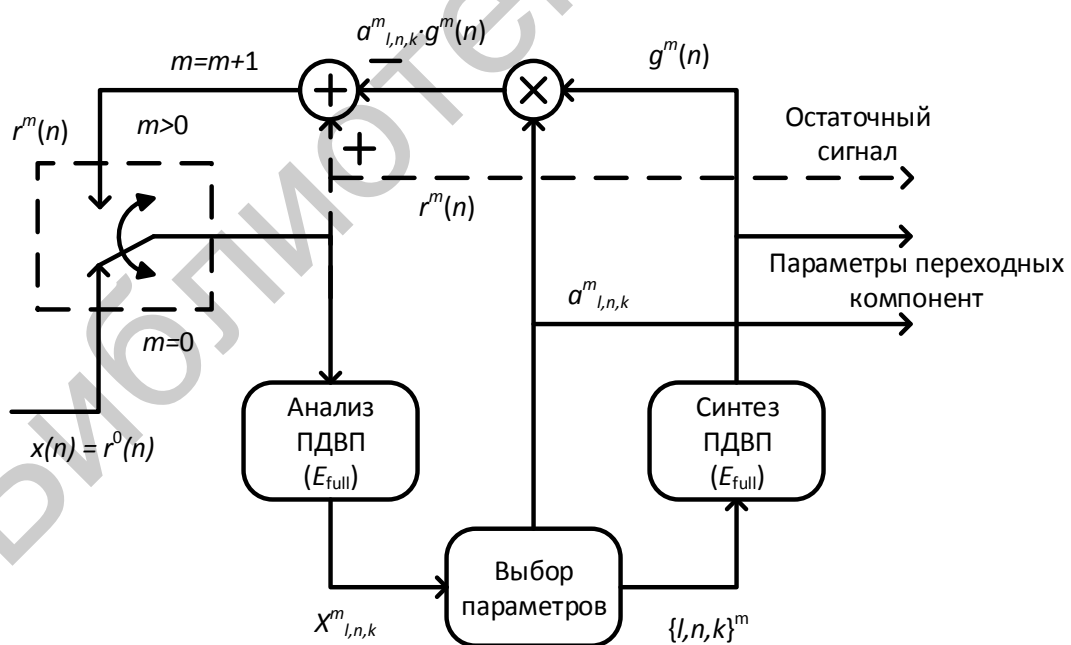


Рисунок 3.3 – Структура алгоритма МР со словарем, построенным на основе полного дерева ПДВП

Процесс повторяется в том случае, когда для последующей итерации алгоритма входным сигналом будет остаточный сигнал с предыдущей итерации. На  $m$ -й итерации остаточный сигнал  $r^m(n)$  вычисляется как

$$r^m(n) = \begin{cases} x(n) & m = 0, \\ r^{m+1}(n) + a_\gamma^m \cdot g_\gamma^m(n) & m \neq 0, \end{cases} \quad (3.7)$$

где  $a_\gamma^m$  – весовой коэффициент оптимального вектора  $g_\gamma^m(n)$  на  $m$ -й итерации;  $\gamma$  – индекс словаря  $D$  на  $m$ -й итерации.

Оптимальным вектором считается тот, у которого получается наибольшее значение произведения с остаточным сигналом  $\langle r^m, g_\gamma^m \rangle$ .

Моделирование работы алгоритма МР выполняется над фреймом входного сигнала, показанного на рисунке 3.4. В основе алгоритма анализа/синтеза фрейма входного сигнала используется полное дерево ПДВП с числом уровней декомпозиции 8 и функция Db20 из семейства Добеши с 40 коэффициентами. Подбор аппроксимирующих функций выполняется по максимуму значений вейвлет-коэффициентов после анализа деревом ПДВП.

На рисунке 3.5 показан результат моделирования алгоритма МР с полным словарем вейвлет-функций ПДВП для входного сигнала, показанного на рисунке 3.4, здесь низкочастотные компоненты выбираются в первую очередь, однако они перцептуально незначимы, что увеличивает число итераций.

Алгоритм МР требует значительных вычислительных затрат от итерации к итерации при формировании субоптимального решения, которое в конце концов может быть неприменимо в некоторых приложениях обработки звука и речи. Задача поиска оптимального словаря  $D$ , построенного на основе ПДВП, сводится к поиску лучшей структуры дерева декомпозиции ПДВП.

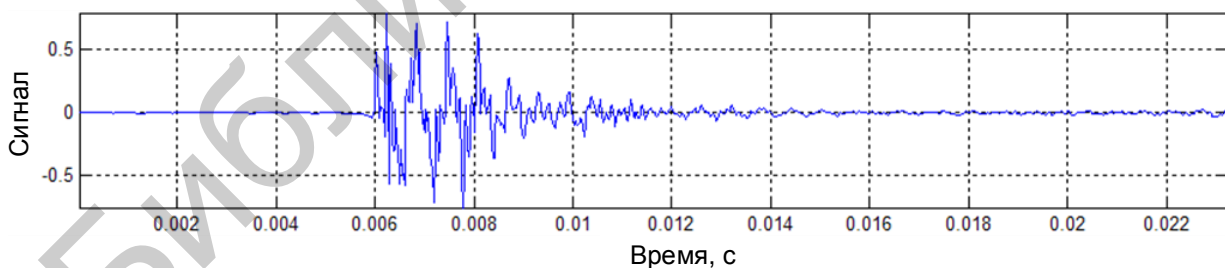
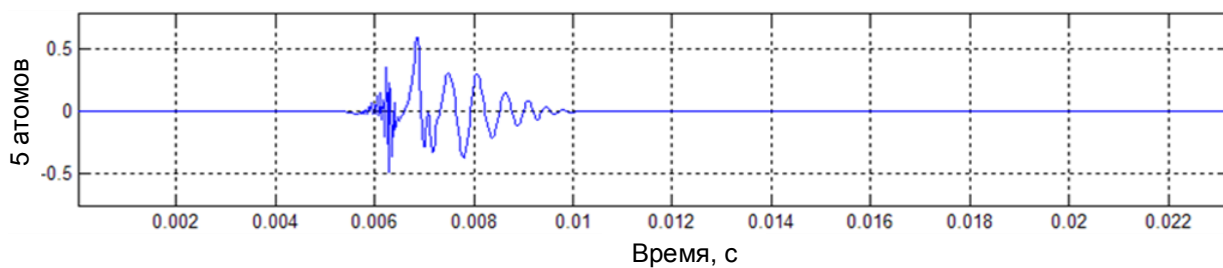
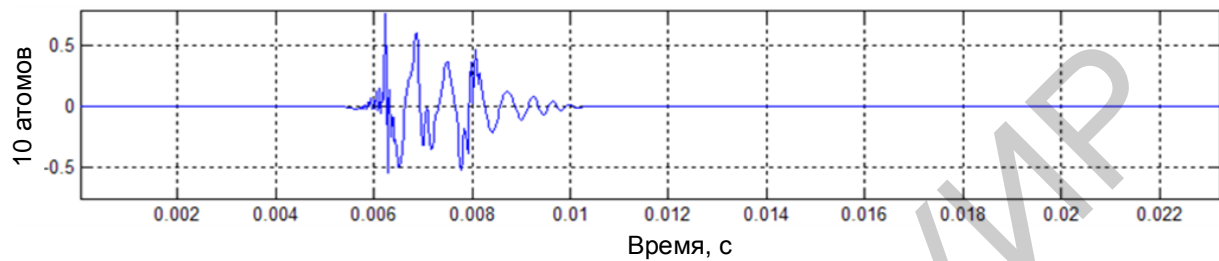


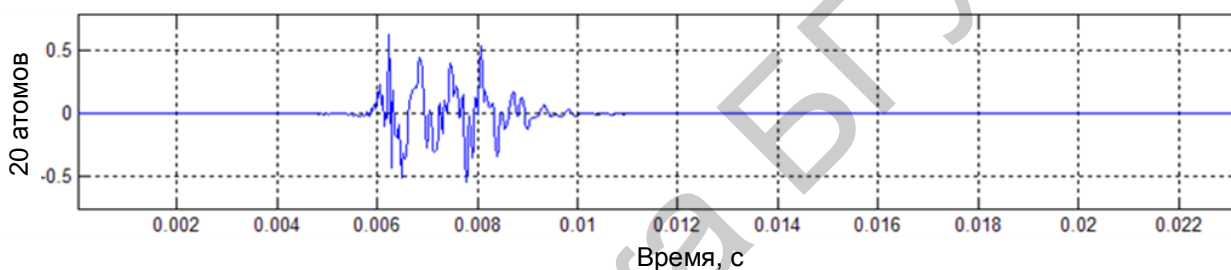
Рисунок 3.4 – Исходный сигнал для параметрического анализа



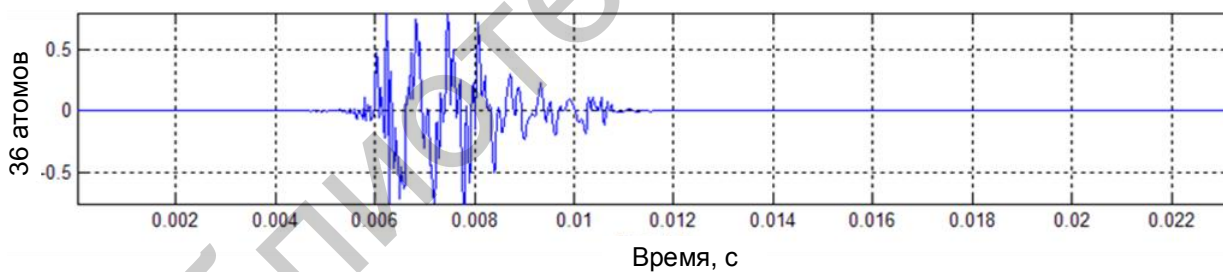
а



б



в



г

а – 5 атомарных функций; б – 10 атомарных функций; в – 20 атомарных функций;  
г – 36 атомарных функций

Рисунок 3.5 – Моделирование алгоритма МР с полным словарем вейвлет-функций ПДВП

### 3.3 Разреженная аппроксимация аудиосигналов с перцептуально-оптимизированным ПДВП-словарем вейвлет-коэффициентов

#### 3.3.1 Схема параметрического анализа

Задача параметрического анализа аудиосигналов решается на основе применения алгоритма МР, где формирование оптимального набора атомов – частотно-временных функций в словаре  $D$  к входному сигналу – выполняется на основе применения перцептуально-адаптированного к фрейму входного сигнала ПДВП. Применение перцептуального критерия позволит уменьшить размер формируемого словаря, а формирование словаря от фрейма к фрейму сделает его динамическим, т. е. зависимым от структуры входного сигнала.

Структурная схема параметрического анализа на основе ПДВП включает две части: в первой части располагается схема оптимизации структуры дерева ПДВП под фрейм входного сигнала на основе стоимостных функций. Оптимизация структуры дерева ПДВП производится согласно описанному алгоритму анализа входного сигнала на основе адаптивного ПДВП, как это было описано в подразделе 3.2. Структурная схема, реализующая алгоритм нахождения оптимальной декомпозиции для входного сигнала на основе адаптивного ПДВП, показана на рисунке 3.6.

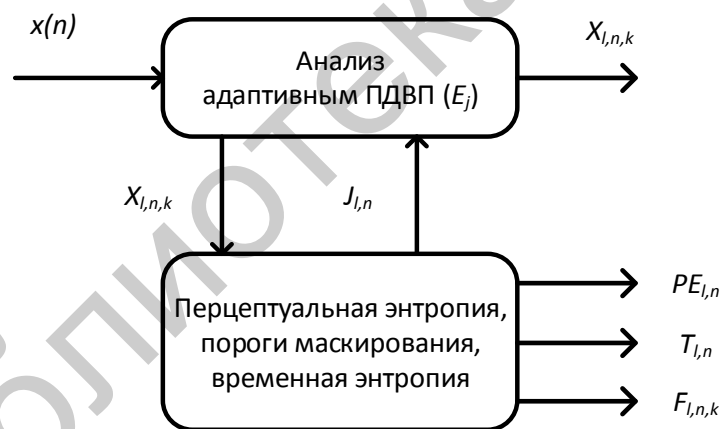


Рисунок 3.6 – Структура декомпозиции входного сигнала при использовании адаптивного ПДВП

Во второй части схемы выполняется модифицированный алгоритм МР с перцептуально-оптимизированным словарем, построенным на основе адаптированного дерева ПДВП в первой части схемы. Данный подход позволит уменьшить число итераций алгоритма МР и производить выбор только значимых компонент сигнала, что будет соответствовать оптимальным векторам декомпозиции. Однако поиск наилучшей аппроксимации анализируемого входного сигнала

$x(n)$  векторами  $g_{l,n,k}$  из словаря  $D$ , построенного на основе перцептуально-адаптированного ПДВП, – сложная в вычислительном плане задача. Количество атомов словаря  $D$  находится в прямой зависимости от числа уровней дерева ПДВП и длины анализируемого фрейма входного сигнала. С точки зрения, например, перцептуального восприятия акустической информации человеком наиболее критичные компоненты сигнала расположены в области нижних, нежели в области верхних частот, что сказывается на качестве выбора перцептуально-значимых полос для анализа.

Минимизация вычислительной сложности связана с минимизацией словаря  $D$ , т. е. сокращение числа доступных вариантов перебора для текущего фрейма обрабатываемого сигнала. Такая минимизация реализуется в модифицированном алгоритме МР, где в алгоритм МР введена процедура расчета перцептуальной энтропии  $PE_{l,n}$ . В этом случае итерационный пересчет значения перцептуальной энтропии  $PE_{l,n}^m$  производится для каждой полосы  $(l, n)$ , на основании которой ведется селекция перцептуально-значимых частотных полос. Знание не только частотной составляющей сигнала, т. е. порогов маскирования  $T_{l,n}$ , но и временного маскера  $F_{l,n,k}$ , позволяет обеспечить оптимальный выбор значимых компонент сигнала.

Структура модифицированного алгоритма МР с динамически оптимизированным словарем показана на рисунке 3.7, где на итерации  $m = 0$  переключатели SW1 и SW2 включены так, чтобы остаточный сигнал  $r^m(n)$  стал равным фрейму входного сигнала  $x(n)$ , а вейвлет-коэффициенты  $X_{l,n,k}^0$  попали напрямую в блок скалярного произведения и скалограммы возбуждения соответственно. Первоначальное значение перцептуальной энтропии  $PE_{l,n}^0$ , порогов маскирования  $T_{l,n}^0$  и временных маскеров  $F_{l,n,k}^0$  берется из схемы декомпозиции входного сигнала адаптивным ПДВП. В блоке скалярного произведения и скалограммы возбуждения выполняется формирование скалограммы возбуждения для текущего входного фрейма сигнала  $G_{l,n,k}^0$ , отбор по ней перцептуально значимого вейвлет-коэффициента, формирование весового параметра атома  $\alpha_{l,n,k}^0$  и позиции вейвлет-коэффициента. В блоке обратного ПДВП выполняется синтез единичной атомарной функции  $g^0(n)$  на основании знания позиции вейвлет-коэффициента в структуре дерева  $E_j$  ПДВП, затем умножение ее на весовой коэффициент  $\alpha_{l,n,k}^0$ . Остаточный сигнала, он же входной сигнал для следующей итерации, формируется вычитанием результирующего сигнала  $\langle \alpha_{l,n,k}^0 \cdot g^0(n) \rangle$  из сигнала  $r^0(n)$ .

На каждой последующей итерации  $m > 0$  пороги маскирования  $T_{l,n}^m$  и временные маскеры  $F_{l,n,k}^m$  вычисляются для формирования скалограммы возбуждения  $G_{l,n,k}^m$ , на основании которой ведется отбор перцептуально-значимых компонент сигнала.

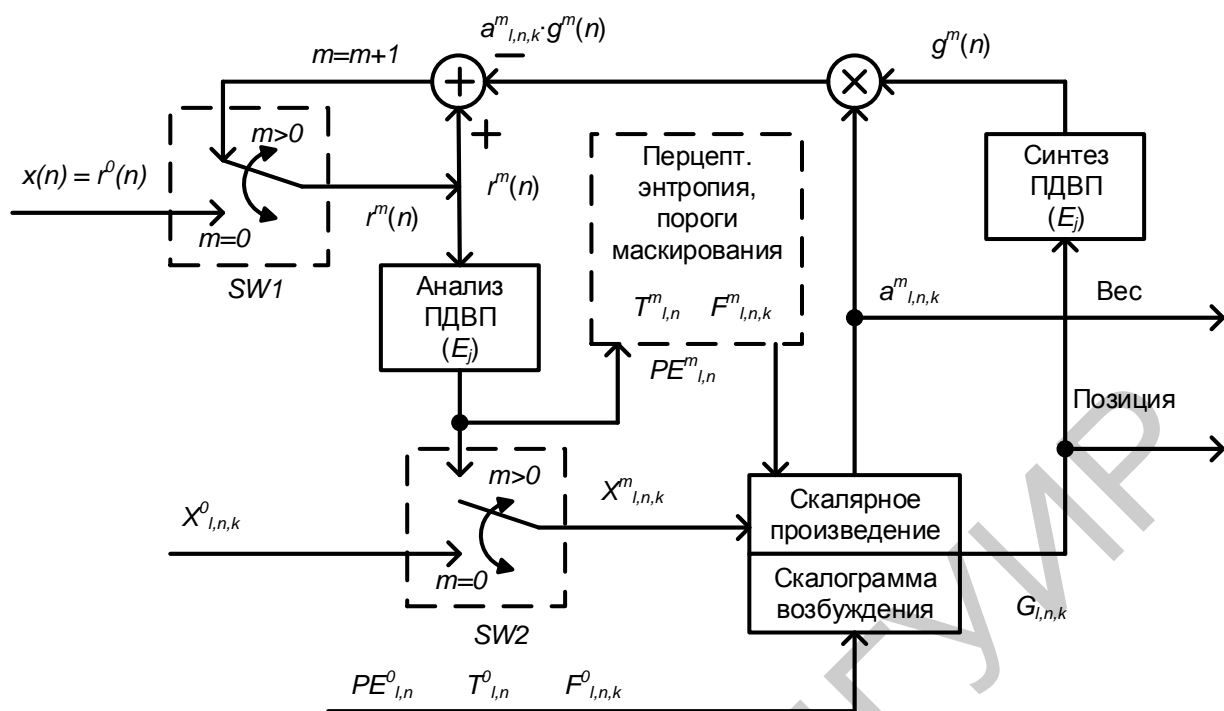


Рисунок 3.7 – Схема параметрического анализа аудиосигналов на основе разреженной аппроксимации с перцептуально-оптимизируемым ПДВП-словарем

### 3.3.2 Процедура расчета порогов маскирования в вейвлет-область

Пусть дано  $(l, n) \in E_j$ , частотно-временной план дерева ПДВП  $E_j$  и коэффициенты  $X_{l,n,k}$ . Процедура расчета порогов маскирования следующая :

- 1) Вычислить спектральную энергию субполосы  $z = (l, n)$ :

$$A(z) = \sum_{k=0}^{K-1} X_{z,k}^2 \quad (3.8)$$

- 2) Оценить тональность сигнала в каждой частотной субполосе и значения индексов  $a_{tmn}(z)$  и  $a_{nmn}(z)$  уменьшения спектральной энергии барка соответственно для тоновых и шумовых маскеров.

- 3) Индекс  $a_{tmn}(z)$ , который оценивает отношение маскирования тоном шума, задается так:

$$a_{tmn}(z) = -0,275 \cdot z - 15,025 \text{ [дБ]}. \quad (3.9)$$

- 4) Индекс маскирования шумом шума  $a_{nmn}(z)$  оценивается как константа

$$a_{nmn}(z) = -25 \text{ [дБ]}, \quad (3.10)$$

так как ПДВП уже внесло некоторое спектральное перекрытие.



5) Среднее значение тональности маскеров в каждой субполосе определяется маскирующим индексом:

$$a_{CB}(z) = \eta \cdot a_{tmn}(z) + (1 - \eta) \cdot a_{nmn}(z) [\text{дБ}], \quad (3.11)$$

где  $\eta$  – тональный коэффициент:

$$\eta = \min(SFM(z) / SFM_{\max}, 1), \quad (3.12)$$

где  $SFM_{\max}$  – мера спектральной пологости [дБ], рассчитываемая как

$$SFM(z) = \frac{(\prod_{k=0}^{K-1} X_{z,k}^2)^{\frac{1}{K}}}{\frac{1}{K} \sum_{k=0}^{K-1} X_{z,k}^2}, \quad (3.13)$$

где  $SFM_{\max}$  – максимальное значение меры пологости спектра [дБ]. Для заданного фильтра прототипа  $SFM_{\max} = -25$  дБ.

6) Найти спектральную энергию субполосы  $z$  с учетом тональности сигнала:

$$D_{CB}(z) = 10 \cdot \log \left( A_{CB}(z) \cdot 10^{\frac{a_{CB}(z)}{10}} \right) [\text{дБ}]. \quad (3.14)$$

7) Вычислить разброс энергии субполосы  $C_{CB}(z)$  как свертку  $D_{CB}(z)$  с функцией разброса  $B(z)$  в каждой частотной субполосе  $z$ :

$$C_{CB}(z) = 10 \cdot \log \left( \frac{1}{K} \sum_{k=1}^{25} 10^{\frac{D_{CB}(k)}{10}} \cdot 10^{\frac{B(z-k)}{10}} \right) [\text{дБ}], \quad (3.15)$$

где  $z = 1, \dots, 25$ , а функция разброса  $B(z)$  вычисляется как

$$B(z) = a + \frac{v+u}{2} \cdot (z+c) - \frac{v-u}{2} \cdot \sqrt{d+(z+c)^2}, \quad (3.16)$$

а параметры функции  $a, v, u, d, c$  приведены в первой строке таблицы 3.1.

8) Вычислить пороги маскирования во временной области:

Временное маскирование в перцептуальной модели восприятия звука человеком основано на манипулировании двумя эффектами маскирования: маскирование вперед (pre-masking) и маскирование назад (post-masking). Временное маскирование назад более важно, чем маскирование вперед, так как маскер оказывает более длительное воздействие в интервале до 200 мс. ПДВП выполняет

разложение сигнала на две области (масштабную и частотную). Расчет временного маскирования выполняется над вейвлет-коэффициентами в каждой полосе  $(l, n)$ :

– вычислить энергию вейвлет-коэффициентов в каждой частотной субполосе  $z$ :

$$E_z(k) = X_{z,k}^2, k = \overline{1, K-1}; \quad (3.17)$$

– определить временную функцию разброса энергии в каждой частотной субполосе  $z$  как свертку  $E_z(k)$  и функции разброса  $B(k)$ :

$$F_z(m) = \frac{1}{K} \sum_{k=0}^{K-1} E_z(k) \cdot 10^{\frac{B(K-k)}{10}}, m = \overline{1, K-1}; \quad (3.18)$$

– найти временной фактор маскирования в субполосе  $z$  как результат сравнения величин:

$$F_z(m) \geq E_z(m), k = \overline{1, K-1}. \quad (3.19)$$

Если данное соотношение выполняется, то в соответствующей частотной субполосе имеет место временное маскирование, в противном случае его нет.

Таблица 3.1 – Параметры функции разброса

Функция разброса	$v$	$u$	$d$	$c$	$a$
Шкала Барк	30 дБ/барк	-25 дБ/барк	0,3	0,05	15
Временная шкала	0,0825 дБ/ $F_{\min}^*$	-0,0412 дБ/ $F_{\min}^*$	0,3	0,157	0,032/ $F_{\min}^*$
$F_{\min}^*$ – минимальная длина анализируемого фрейма					

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 частотной субполосе  $z$  (см. рисунок 3.8). Максимальное временное разрешение для ПДВП имеет место в частотных субполосах 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

$$B(k) = a + \frac{v+u}{2} \cdot (k+c) - \frac{v-u}{2} \cdot \sqrt{d+(k+c)^2} \text{ [дБ]} \quad (3.20)$$

параметры которой вдоль оси времени определяются как  $v = 20$  дБ/мс =  $= 0,0825$  дБ/ $F_{\min}^*$  и  $u = 20$  дБ/мс =  $-0,0412$  дБ/ $F_{\min}^*$  (см. таблицу 3.1, строка 2). На рисунке 3.8 схематически показано проявление временного маскирования.

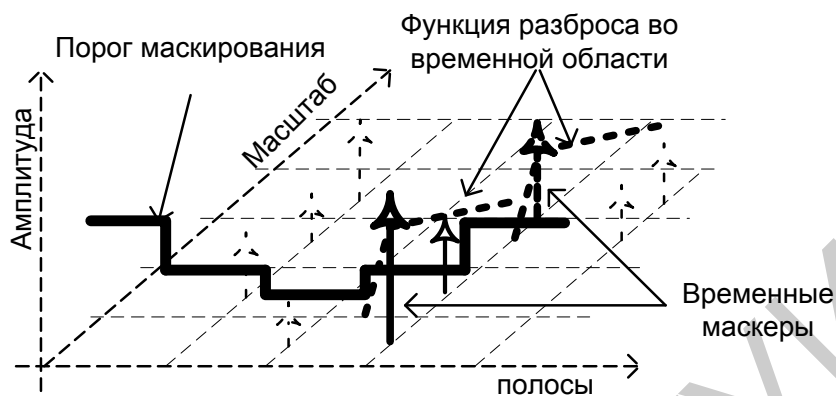


Рисунок 3.8 – Схема временного маскирования в соответствующей частотной субполосе  $z$

9) Оценить частотно-временной маскирующий порог  $M_{CB}(z)$  в каждой частотной субполосе:

$$M_{CB}(z) = C_{CB}(z) \cdot \max\left(\frac{F_z(k)}{E_z(k)}, 1\right) [\text{дБ}], k = \overline{1, K-1}.$$

10) Вычислить значение маскирующего порога  $T_{CB}(z)$  в соответствующей частотной субполосе из сравнения частотно-временного маскирующего порога  $M_{CB}(z)$  с минимальным значением абсолютного порога слышимости  $ATH(z)$ :

$$T_{CB}(z) = \max(ATH(z), M_{CB}(z)) [\text{дБ}]. \quad (3.21)$$

Рассчитанные пороги маскирования  $T_{l,n}$  в частотной области, в соответствии с алгоритмом, могут быть расширены, если текущая полоса содержит временной маскер, для построения скалограммы возбуждения, которая зависит от двух компонент, одна из которых частота, другая масштаб.

Входной сигнал (см. рисунок 3.4) и соответствующая ему скалограмма возбуждения, где перцептуально-значимые компоненты показаны как положительные элементы, иллюстрируется на рисунке 3.9. Функция Db20 из семейства Добеши с 40 коэффициентами использовалась для анализа входного сигнала в схеме перцептуально-адаптивного ПДВП.

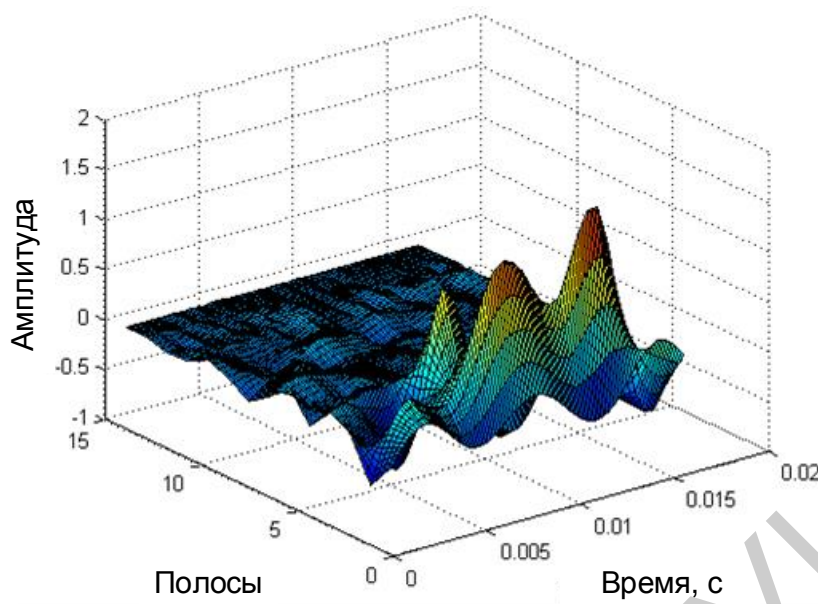


Рисунок 3.9 – Часть скалограммы возбуждения, соответствующей входному сигналу, представленному на рисунке 3.4

### 3.3.3 Алгоритм параметрического анализа аудиосигналов на основе разреженной аппроксимации с перцептуально-оптимизированным словарем

Алгоритм МР с динамически оптимизируемым словарем на основе психоакустического критерия может быть реализован так, как показано ниже.

Исходные данные МР подгонки:  $E_j$  – оптимизированная структура ПДВП для кадра входного сигнала  $x(n)$ ;  $T_{l,n}$  – порог маскирования для каждого оконечного узла  $(l, n) \in E_j$  ПДВП;  $F_{l,n}$  – временный маскер для каждого оконечного узла  $(l, n) \in E_j$  ПДВП;  $G_{l,n,k}$  – скалограмма возбуждения соответствующая кадру входного сигнала  $x(n)$ .

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  $PE_{l,n}^m = 0 \forall (l, n, k) \in E_j$ ,

ТОГДА СТОП;

ЕСЛИ  $PE_{l,n}^m = 0$ ,

ТОГДА  $X_{l,n,k}^m = 0$  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 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  $\alpha_{l,n,k}^m = X_{l,n,k}^{*m}$  (таблица 3.2), 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 (3.7).

Шаг 11. Применить оптимизированный на основе текущего фрейма вейвлет ПДВП 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.

Таблица 3.2 – Значения отобранных функций на каждой  $m$ -й итерации работы алгоритма МР с перцептуально-оптимизированным словарем ПДВП

$m$	$\alpha_\gamma^m$	$l$	$n$	$k$	$m$	$\alpha_\gamma^m$	$l$	$n$	$k$	$m$	$\alpha_\gamma^m$	$l$	$n$	$k$
1	1,7318	5	1	7	13	-0,3160	5	3	6	25	0,1202	4	7	6
2	1,3048	5	1	5	14	-0,3077	5	2	2	26	-0,1148	4	2	1
3	1,1899	5	1	4	15	-0,2983	5	3	5	27	0,1130	5	3	1
4	1,0691	5	3	4	16	-0,2257	5	1	8	28	0,1110	4	2	6
5	0,9415	5	1	6	17	-0,2026	4	2	9	29	0,1021	4	6	15
6	-0,5655	5	3	3	18	-0,1921	4	2	8	30	-0,1436	5	0	1
7	0,5157	5	2	7	19	0,1690	5	1	3	31	0,3806	5	0	8
8	-0,4198	5	1	1	20	-0,1621	4	6	1	32	0,1919	5	0	2
9	0,4132	5	3	7	21	0,1571	5	2	4	33	-0,1774	5	0	6
10	-0,3841	5	2	5	22	-0,1522	4	5	10	34	0,1691	5	0	5
11	-0,3728	5	2	8	23	-0,1453	4	7	9	35	-0,1489	5	0	3
12	0,3399	5	1	2	24	0,1266	4	5	14	36	-0,1017	5	0	7

Преимуществом алгоритма является минимизация искажений в аудиосигнале воспринимаемых на слух благодаря перцептуально-оптимизированному частотно-временному плану соответствующей декомпозиции ПДВП для выбора оптимального атома на каждой итерации подгонки. Более того, существует точный психоакустический критерий остановки описанной процедуры. Число итераций алгоритма МР для анализируемого фрейма определяется числом перцептуально важных вейвлет-коэффициентов ПДВП.

На рисунке 3.10 показан результат моделирования алгоритма МР с перцептуально-адаптированным словарем ПДВП для входного сигнала (см. рисунок 3.4).

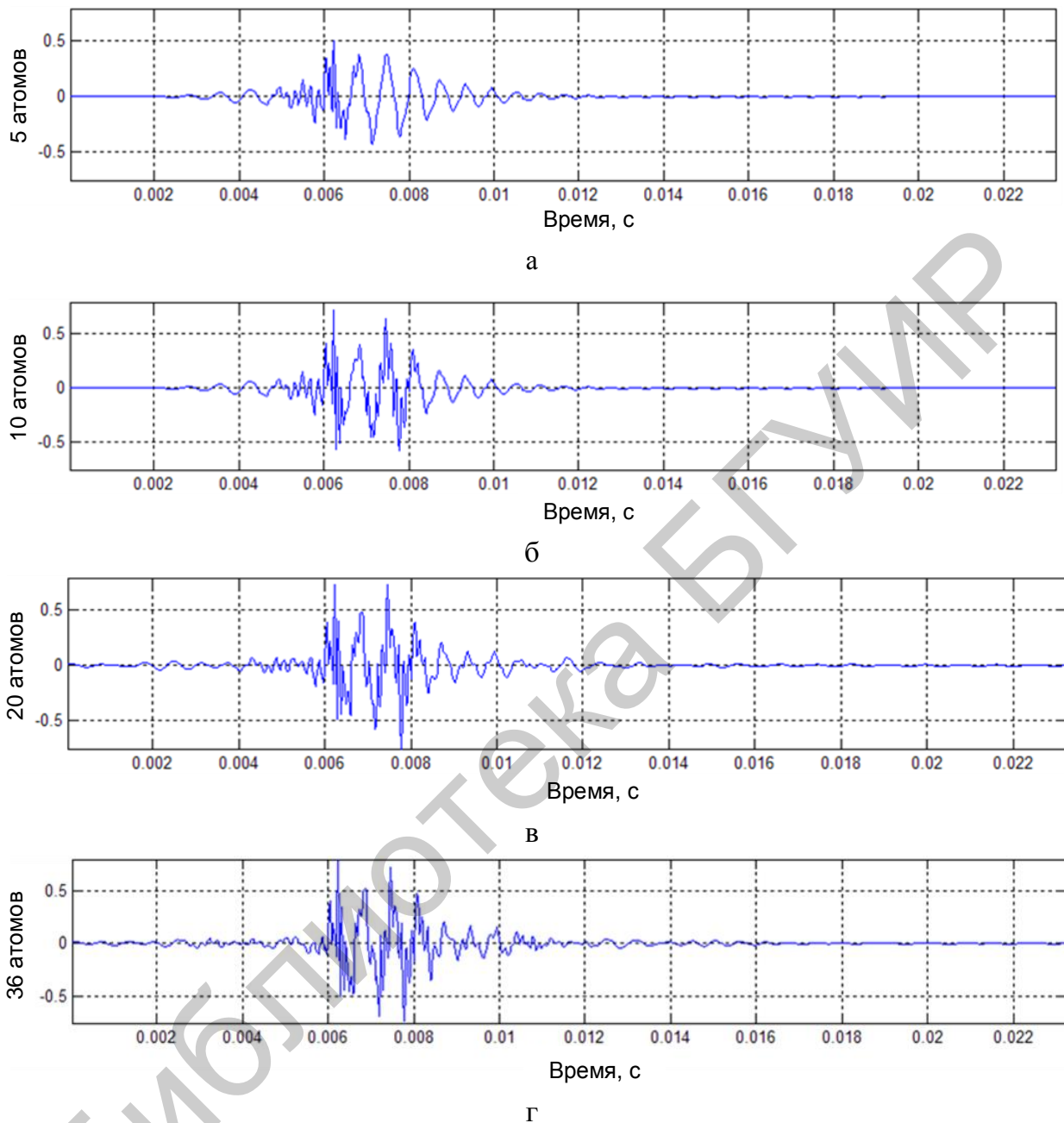


Рисунок 3.10 – Моделирование алгоритма согласованной подгонки с перцептуально-оптимизированным словарем вейвлет-функций ПДВП

Входной сигнал хорошо аппроксимируется небольшим числом атомов (36-ю перцептуально-оптимизированными функциями, см. таблицу 3.2). Это потому, что не требуется сглаживания острых краев. Для моделирования использовалась функция Db20 из семейства Добеши с 40 коэффициентами. Структура адаптированного дерева ПДВП показана на рисунке 3.11 и ассоциируемая с ним АЧХ субполосного банка фильтров приведена на рисунке 3.12.

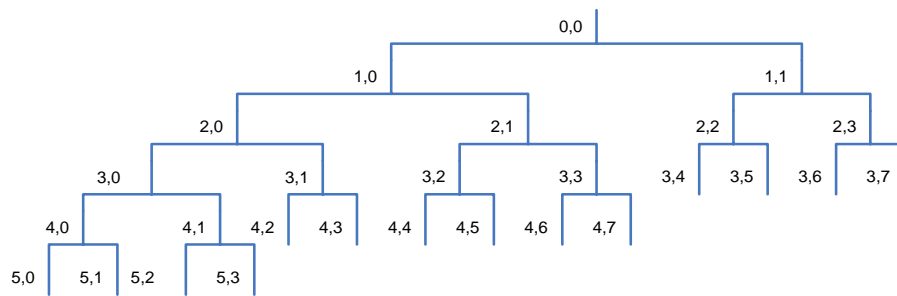


Рисунок 3.11 – Структура дерева ПДВП, адаптированного на основе перцептуального критерия

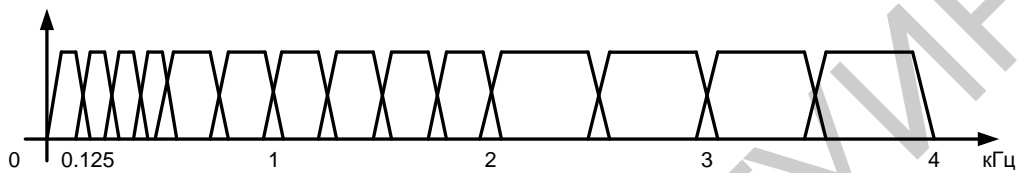


Рисунок 3.12 – АЧХ субполосного банка фильтров, ассоциируемого со структурой дерева ПДВП, показанного на рисунке 3.11

На рисунке 3.13 показана скалограмма возбуждения, ассоциируемая с сигналом (рисунок 3.10, а) при использовании пяти перцептуально значимых атомов для моделируемого сигнала (см. рисунок 3.4). Спектрограмма и порог маскирования для входного и остаточного сигналов после параметрического анализа 36-ю функциями показаны на рисунке 3.14.

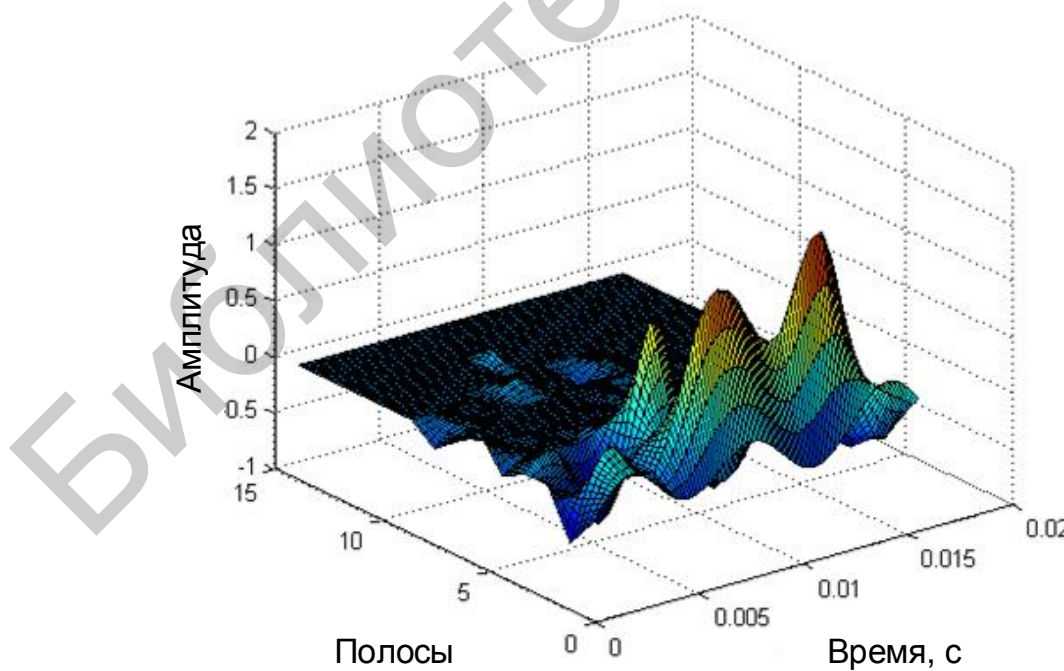
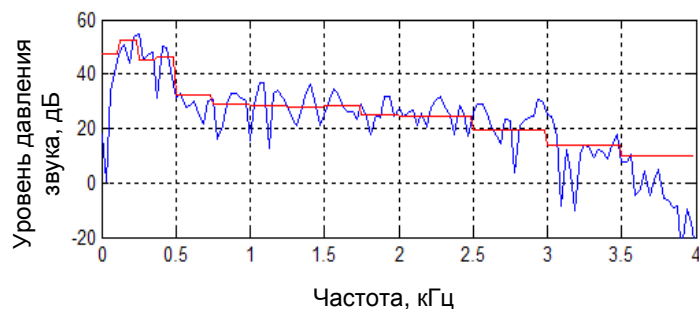
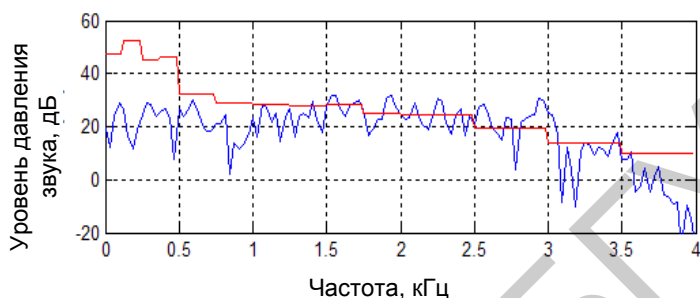


Рисунок 3.13 – Часть скалограммы возбуждения, ассоциируемой с аппроксимируемым сигналом, на основании пяти перцептуально значимых атомов



а



б

Рисунок 3.14 – Спектрограмма и порог маскирования для входного сигнала (а) и для остаточного сигнала после аппроксимации 36-ю перцептуально-оптимизированными функциями (б)

Применение перцептуально-адаптированной скалограммы в алгоритме согласованной подгонки для селекции компонент позволяет определить распределение компонент не только в частотной области, но и во временной. При этом при моделировании будут отобраны только перцептуально значимые компоненты (вейвлет-коэффициенты), а замаскированные и компоненты шума останутся «спрятанными» под порогом и не будут участвовать в отборе.

Результаты оценок объективного тестирования качества аппроксимированного сигнала различным числом функций представлены в таблице 3.3, где отрицательные значения  $NMR$  показывают нижнюю границу порога восприятия, а положительные значения  $SNR$  – энергетическую оценку звуковых искажений. Однако с увеличением числа итераций объективные оценки значительно улучшаются.

Таблица 3.3 – Объективные оценки качества аппроксимированного сигнала в зависимости от числа итераций  $m$

$m$	$NMR_{Total}$ , дБ	$NMR_{Seg}$ , дБ	$SNR_{Total}$ , дБ	$SNR_{Seg}$ , дБ	$MBSD$
3	-0,74	1,13	-3,36	-0,17	4,35
10	-1,77	0,61	-2,56	0,22	1,89
25	-2,29	0,35	-2,14	0,43	1,85
36	-3,8	-0,44	-0,63	1,18	1,25



Сравнение сходимости алгоритма МР на основе трех различных словарей аппроксимирующих функций показано на рисунке 3.15. Алгоритм МР со словарем перцептуально-оптимизированных функций ПДВП обеспечивает наименьшее среднеквадратичное отклонение по сравнению с алгоритмом МР на основе словаря с полным набором вейвлет-функций ПДВП или затухающих синусоид.

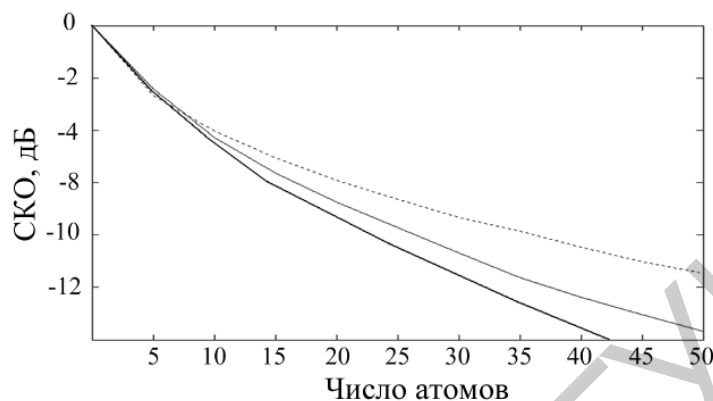


Рисунок 3.15 – Среднеквадратичное отклонение: затухающие синусоиды (пунктирная линия); вейвлет-пакет с избыточным словарем (тонкая сплошная линия); вейвлет-пакет, оптимизированный для каждого фрейма на основе психоакустики (толстая сплошная линия)

### 3.4 Параметрический аудиоречевой кодер на основе разреженной аппроксимации с перцептуально-оптимизированным словарем частотно-временных функций

#### 3.4.1 Структура параметрического аудиокодера

Обобщенная структура параметрического перцептуального аудиокодера показана на рисунке 3.16, где основу составляют адаптивный ПДВП (отмечен пунктирной линией) и часть подбора параметров (серая область, выделенная пунктирной линией).

Анализирующий банк фильтров построен на адаптивном ПДВП, где, как и в структуре перцептуального кодера аудиосигналов, для каждого входного фрейма выполняется подбор оптимальной структуры дерева  $E_i$ , соответствующей наилучшему частотно-временному плану вейвлет-анализа. Блок «Split» рассчитывает параметры реконфигурации дерева ПДВП  $r_{l,n}$  с учетом значений перцептуальной энтропии  $PE_{l,n}$  и энтропии  $H_{E_i}$  структуры дерева ПДВП  $E_i$ . Результатом адаптивного анализа являются вейвлет-коэффициенты  $X_{l,n,k}$ , трехмерная скалограмма возбуждения обрабатываемого фрейма сигнала, построенная на их основе и адаптированная структура дерева ПДВП  $E_j$ .

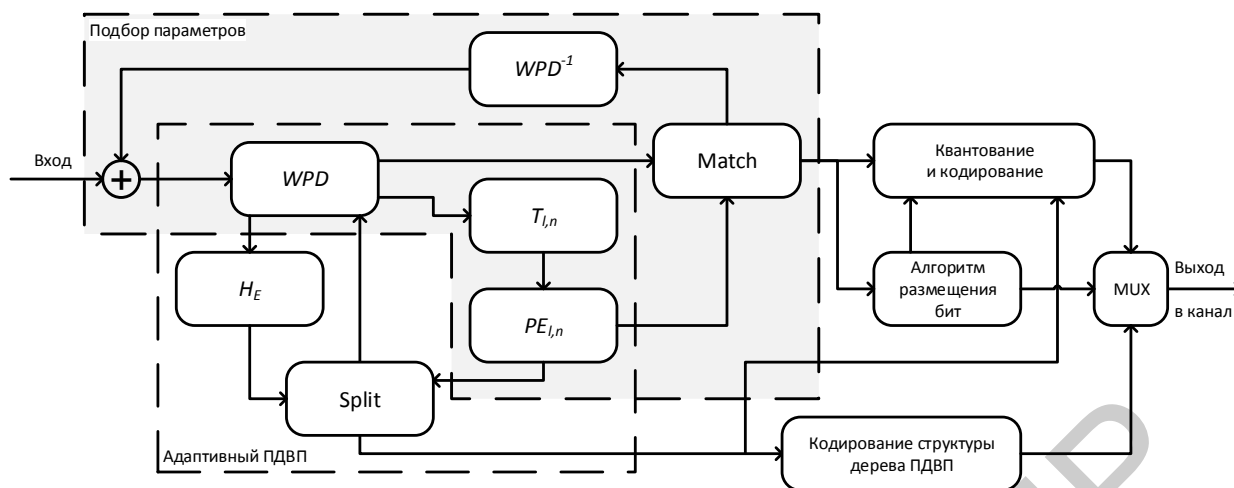


Рисунок 3.16 – Структура параметрического кодера звука

Часть подбора параметров базируется на прямом и обратном вейвлет-преобразовании, блоках расчета порогов маскирования  $T_{l,n}$  и формирования перцептуальной энтропии  $PE_{l,n}$  для каждого узла дерева преобразования и блоке «Match», который отбирает компоненты и формирует параметры. Процесс формирования параметров основан на текущей скалограмме обрабатываемого фрейма, где наиболее перцептуально значимая компонента сигнала с учетом его вклада в трехмерную скалограмму возбуждения отбирается. На основании отобранных компонент выполняется синтез временной функции в блоке « $WPD^{-1}$ », которая далее вычитается из сигнала. Остаточный сигнал анализируется в блоке « $WPD$ » и процесс повторяется. Отобранные компоненты квантуются и передаются в блок кодирования.

Структура закодированных перцептуально значимых компонент такова, что на стороне декодера после деквантования параметры записываются в уже подготовленную структуру дерева ПДВП и выполняется синтез сигнала. Структура ПДВП-декодера аудиосигналов показана на рисунке 3.17. Работа декодера выполняется в следующем порядке: входная информация декодируется и восстанавливается в блоке декодирования и восстановления. Полученные параметры содержат значение весового коэффициента и информацию о размещении декодированного параметра в структуре дерева ПДВП. Заполнение параметрами структуры дерева преобразования происходит в блоке размещения коэффициентов. Реконструкция аудиосигнала выполняется в блоке « $WPD^{-1}$ » на основе обратного ПДВП в соответствии со структурой полного дерева ПДВП. Синтезированные кадры реконструированного сигнала умножаются на модифицированное треугольное окно и складываются для формирования выходной последовательности  $\hat{x}(n)$ .

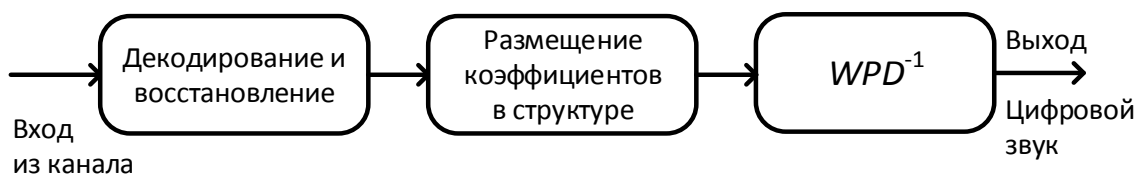


Рисунок 3.17 – Структура параметрического декодера аудиосигналов на основе полного дерева ПДВП

### 3.4.2 Экспериментальные результаты и оценка реконструированного аудиосигнала

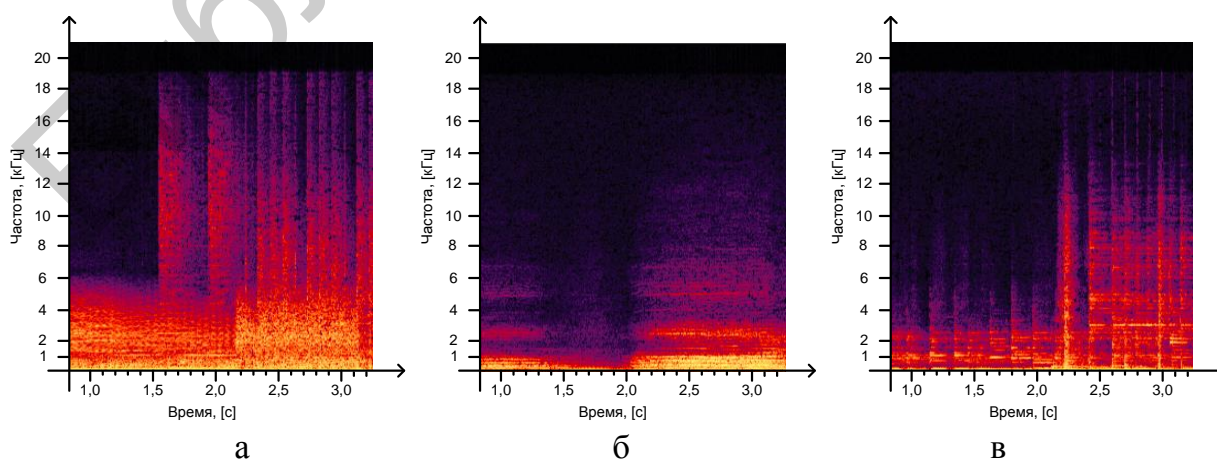
Результаты экспериментальных исследований параметрического аудиокодера на основе разреженной аппроксимации с перцептуально-оптимизированным словарем частотно-временных функций сведены в таблицу 3.4.

Таблица 3.4 – Оценки реконструированного сигнала параметрическим аудиокодером

Тестовый материал	MP3			AAC			Параметрический кодер		
	Степень сжатия	Субъективное отличие	MBSD	Степень сжатия	Субъективное отличие	MBSD	Степень сжатия	Субъективное отличие	MBSD
Rock	7,7	−0,2	1,45	9,5	−0,067	1,17	18,3	−0,3	1,55
Classic	16,6	−0,067	1,02	9,6	−0,2	1,32	20,2	−0,4	1,67
Pop	8,4	−0,045	1,13	11,5	−0,067	1,16	14,7	−0,2	1,34

В таблице 3.4 показаны сравнительные оценки степени компрессии, субъективного и объективного характера для разных аудиосигналов (Rock, Classic, Pop). Тестовые исходные сигналы Rock, Classic, Pop оцифрованы с частотой дискретизации 44,1 кГц и представлены 16-битным кодом ИКМ. Спектрограммы реконструированных аудиосигналов показаны на рисунке 3.18, а–в соответственно.

Оценка степени компрессии получена для сигналов с почти равным субъективным качеством реконструированного сигнала. Как видно из таблицы 3.4, среднее значение степени компрессии для параметрического кодера в 1,7 раз больше по сравнению с другими. В таблице 3.4 представлены объективная оценка *MBSD* для тестовых сигналов, которая хорошо коррелирует с субъективной оценкой *MOS* и значением субъективного отличия, т. е. значение *MBSD*, стремящееся к единице, соответствует субъективной оценке *MOS*, стремящейся к пяти и значению субъективного отличия, стремящегося к нулю.



а – Rock; б – Classic; в – Pop

Рисунок 3.18 – Спектрограммы экспериментальных сигналов

## ЧАСТЬ 2 СОВРЕМЕННЫЕ ПРЕОБРАЗОВАНИЯ ДЛЯ ПРИЛОЖЕНИЙ КОДИРОВАНИЯ ИЗОБРАЖЕНИЙ

### 4 ТРАНСФОРМАЦИОННОЕ КОДИРОВАНИЕ ЦИФРОВЫХ ИЗОБРАЖЕНИЙ

#### 4.1 Система трансформационного кодирования

Цифровое изображение  $\mathbf{x}$ , состоящее, например, из  $N$  строк и  $N$  столбцов как некий тип данных математически можно описать дискретной функцией двух переменных  $f(x, y)$  с областью определения  $\mathbf{E} = \{0, \dots, N-1\} \cdot \{0, \dots, N-1\}$ , называемой пространственной областью, и областью значений  $\mathbf{D} = \{0, 1, \dots, L-1\}$ . Таким образом, цифровое изображение – это прямоугольная матрица  $\mathbf{A} = (a_{ij})_{i=1, j=1}^{N, N}$ ,  $a_{ij} \in \{0, 1, \dots, L-1\}$ ,  $i = \overline{1, N}$ ,  $j = \overline{1, N}$ , элементы которой (пиксели) принимают значения из некоторого отрезка ряда целых чисел. Указанный отрезок начинается с нуля, поскольку в общем случае значение пикселя определяет его яркость и отрицательные значения не рассматриваются. Мощность множества возможных значений пикселя  $L$ , т. е. количество различных уровней яркости, для удобства аппаратной и программной реализации всегда принимается равной некоторой степени двойки, чтобы каждый пиксель мог быть представлен целым числом битов.

В зависимости от значения  $L$ , наличия цвета у пикселей, а также особенностей содержимого различают несколько основных типов цифровых изображений:

1) Монохромные (бинарные), где каждый пиксель изображения кодируется одним битом, например, факсимильная связь.

2) Полутоновые цифровые изображения, когда все множество возможных значений пикселя воспринимается как переход от черного цвета к белому через некоторое количество промежуточных оттенков серого (уровней яркости).

3) Цветные цифровые изображения, которые представляются с помощью определенной цветовой модели:

- непрерывно-тоновые (оцифрованные изображения окружающей действительности: плавные цветовые переходы и отсутствие резких границ);
- дискретно-тоновые (изображения компьютерной графики и мультипликации).

Данную классификацию можно продолжить. Например, своей спецификой обладают рентгеновские снимки, снимки из космоса, изображения, зарегистрированные в инфракрасном диапазоне, и т. д.

Цифровые изображения обладают важнейшей с точки зрения сжатия особенностью: с высокой степенью вероятности соседние пиксели в пределах некоторой локальной области имеют одинаковые или близкие цвета. Это достаточно очевидный факт, и трудно привести пример полезного изображения, у которого

все соседние пиксели будут резко отличаться друг от друга. Избыточность такого типа, присущая лишь изображениям, получила название пространственной. Сжатие изображения заключается в устранении корреляции между пикселями посредством специального преобразования и дальнейшего статистического кодирования преобразованных данных.

Методы кодирования изображений, которые оперируют непосредственно со значениями элементов изображения, например значениями яркостей пикселей, являются пространственными методами. Методы компрессии изображений, основанные на модификации и сжатии результатов преобразования изображения, называются методами трансформационного кодирования. Согласно данному подходу обратимое линейное преобразование  $\mathbf{T}$ , представляющее собой квадратную матрицу, используется для отображения множества целых чисел исходного изображения  $\mathbf{x}$  во множество вещественных чисел коэффициентов  $\mathbf{u}$  преобразования  $\mathbf{T}$ , которые затем квантуются и кодируются.

Система трансформационного кодирования изображений (сжатия – кодер и восстановления – декодер) показана на рисунке 4.1. В кодере реализуются четыре операции: разбиение изображения размером  $N \times N$  на блоки  $n \times n$  пикселей, прямое преобразование  $\mathbf{T}_A$  – преобразование анализа формата  $n \times n$ , квантование и кодирование. Декодер выполняет обратную последовательность операций: декодирование символов, обратное преобразование  $\mathbf{T}_S$  – преобразование синтеза формата  $n \times n$  и объединение блоков в изображение. Базовые функции преобразования анализа – столбцы  $\Phi_i$  матрицы  $\mathbf{T}_A^T$ , в то время как столбцы  $\Phi_i$  матрицы  $\mathbf{T}_S$  являются базовыми функциями преобразования синтеза. Преобразование будет перфективным, или *биортогональным*, если

$$\mathbf{T}_S \mathbf{T}_A = \mathbf{I}.$$

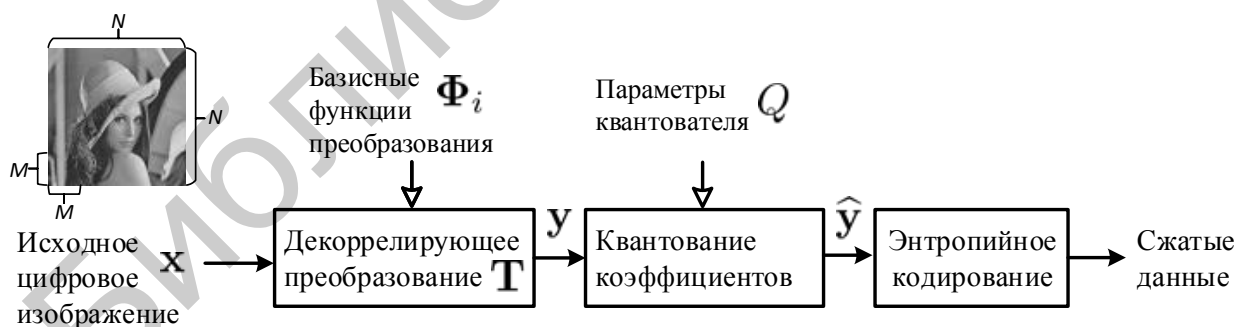


Рисунок 4.1 – Система трансформационного кодирования изображений (сжатия – кодер и восстановления – декодер)

В специальном случае, когда преобразование синтеза имеет действительные коэффициенты, тогда просто выразить преобразование анализа через преобразование синтеза:

$$\mathbf{T}_S = \mathbf{T}_A^{-1} = \mathbf{T}_A^T.$$

Преобразование  $\mathbf{T}_A$  здесь будет *ортонормальным*, или *ортогональным*. Свойство ортогональности преобразования часто используется в системах кодирования изображений, потому что энергия сигнала сохраняется и тем самым не происходит увеличения энергии ошибок (например, из-за квантования) в восстановленном изображении.

Целью операции преобразования является декорреляция значений элементов  $\mathbf{X}$  в каждом блоке или, другими словами, уплотнение как можно большего количества информации в наименьшее число коэффициентов преобразования  $\mathbf{Y}$ . Корреляционная матрица коэффициентов  $\mathbf{Y}$  преобразования  $\mathbf{T}$  равна

$$\mathbf{R}_{yy} = E\{\mathbf{T}\mathbf{x}\mathbf{T}^T\} = \mathbf{T}\mathbf{R}_{xx}\mathbf{T}^T,$$

где  $\mathbf{R}_{xx}$  является корреляционной матрицей исходного изображения  $\mathbf{x}$ ;

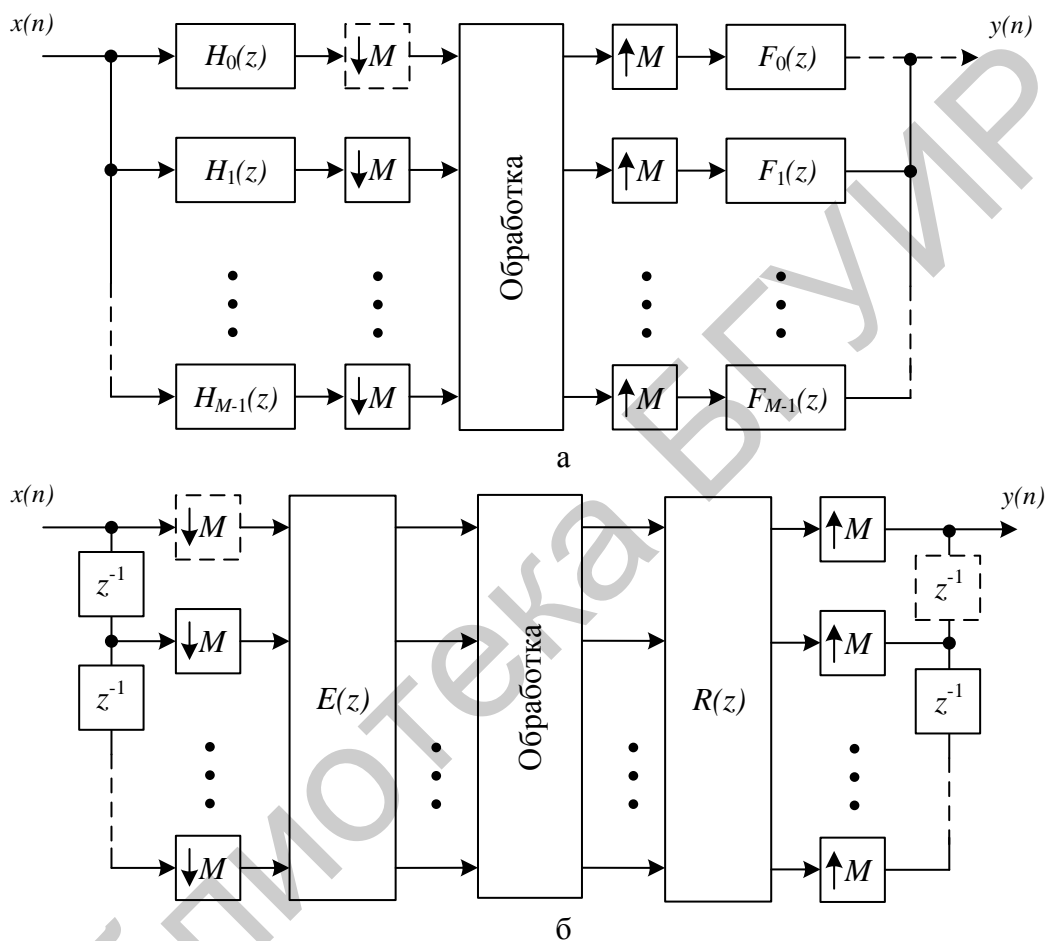
$E$  обозначает операцию математического ожидания.

Здесь линейное преобразование  $\mathbf{T}$  оптимально, если можно полностью декоррелировать значения исходного изображения  $\mathbf{x}$ . Для этого необходимо, чтобы матрица  $\mathbf{R}_{yy}$  была диагональной, у которой элементы, находящиеся на главной диагонали, являются собственными числами матрицы  $\mathbf{R}_{xx}$ , а строки матрицы преобразования  $\mathbf{T}$  представляют собой собственные векторы  $\mathbf{R}_{xx}$ . Причем в первой строке записывается вектор, которому отвечает наибольшее собственное число матрицы  $\mathbf{R}_{xx}$ , а в последней – собственный вектор  $\mathbf{R}_{xx}$ , соответствующий наименьшему собственному числу матрицы  $\mathbf{R}_{xx}$ . Все элементы корреляционной матрицы  $\mathbf{R}_{yy}$ , расположенные вне главной диагонали, равны нулю, поэтому компоненты векторов  $\mathbf{Y}$  являются некоррелированными.

Таким образом, матрицы  $\mathbf{R}_{xx}$  и  $\mathbf{R}_{yy}$  имеют совпадающий набор собственных чисел и векторов. С другой стороны, поскольку строки матрицы  $\mathbf{T}$  представляют собой ортонормированные векторы, отсюда следует, что  $\mathbf{T}^{-1} = \mathbf{T}^T$ , т. е. выполняется условие ортогональности  $\mathbf{T}_S = \mathbf{T}_A^{-1} = \mathbf{T}_A^T$  и, значит, любой вектор  $\mathbf{x}$  может быть восстановлен по соответствующему вектору  $\mathbf{y}$  с помощью соотношения  $\mathbf{x} = \mathbf{T}^T \mathbf{y}$ . Такое преобразование называется преобразованием Карунена – Лоэва (ПКЛ). Однако поскольку ПКЛ зависит от преобразуемых данных, то получение базисных функций  $\Phi_i$  преобразования для каждого блока изображения является нетривиальной вычислительной задачей (высокая вычислительная сложность), а также потребуются большой битовый поток при передаче данных в декодер, если исходное изображение нестационарное, например видеопоток изображений.

Таким образом, стоит задача поиска быстрых, независимых от сигнала преобразований для практических приложений. Один из подходов – это представление сигналов стационарными моделями, поиск аппроксимации ПКЛ и затем разработка быстрых вычислительных алгоритмов для найденных аппроксимаций.

В последние несколько декад банки фильтров рассматриваются как наиболее эффективная техника компрессии мультимедиаданных. Банки фильтров применяются в кодировании аудиосигналов, изображений и видеопотока в таких стандартах, как JPEG, JPEG2000, JPEG XR, MPEG и H.264/AVC. На рисунке 4.2 показана схема максимально децимированного  $M$ -канального банка фильтров, состоящего из систем анализа и синтеза (рисунок 4.2, а – стандартная схема банка), а также банк фильтров из полифазных матриц  $\mathbf{E}(z)$  и  $\mathbf{R}(z)$ .



а – стандартная схема банка фильтров;

б – банк фильтров на основе полифазных матриц  $\mathbf{E}(z)$  и  $\mathbf{R}(z)$

Рисунок 4.2 – Схема максимально децимированного  $M$ -канального банка фильтров

Полифазная структура банка фильтров формулируется следующим образом:

$$[H_0(z) \ H_1(z) \ \dots \ H_{M-1}(z)]^T = \mathbf{E}(z^M)\mathbf{e}(z)^T,$$

$$[F_0(z) \ F_1(z) \ \dots \ F_{M-1}(z)]^T = \mathbf{e}(z)\mathbf{R}(z^M),$$

где  $\mathbf{e}(z) = [1 \ z^{-1} \ \dots \ z^{-(M-1)}]$ ,  $H_k(z)$  и  $F_k(z)$  – передаточные функции  $k$ -го канала банка фильтров анализа и синтеза соответственно.

Импульсная характеристика  $k$ -го канала банка фильтров анализа с позиции стандартного преобразования является базовой функцией преобразования анализа  $\Phi_k$ . Соответственно, импульсную характеристику  $k$ -го канала банка фильтров синтеза можно рассматривать как базовую функцию  $\Phi_k$  преобразования синтеза. Эффективность данной системы заключается в том, что фильтруется не входной сигнал, а сигнал после децимации, который имеет в  $M$  раз меньше отсчетов, что снижает требования к быстродействию каждого вычислителя в  $M$  раз по сравнению с прямой реализацией на одном вычислителе.

Если матрица  $\mathbf{E}(z)$  обратима, то полифазная матрица синтеза  $\mathbf{R}(z)$  может быть выбрана как обратная матрица  $\mathbf{E}(z)$ , тогда совершенная реконструкция достигается

$$\mathbf{R}(z)\mathbf{E}(z) = bz^{-1}\mathbf{I}, b \neq 0, l \geq 0. \quad (4.1)$$

Такой банк фильтров называется банк фильтров с совершенной реконструкцией входного сигнала  $\mathbf{X}$ , или *биортогональный* банк фильтров (БОБФ). Если

$$\mathbf{R}(z) = \mathbf{E}^T(z^{-1}), \quad (4.2)$$

то это специальный класс банка фильтров – *параунитарный* банк фильтров (ПУБФ).

Если длина импульсной характеристики фильтра прототипа больше числа каналов, тогда компоненты данных полифазных матриц являются полиномами переменной  $z$ . Для случая  $M = 2$  двухканальный банк фильтров – это вейвлет-преобразование. Когда длина импульсной характеристики фильтра прототипа равна числу каналов  $M$ , то полифазные матрицы  $\mathbf{E}(z)$  и  $\mathbf{R}(z)$  образованы из полиномов нулевого порядка – скалярных величин и они эквивалентны стандартным преобразованиям, т. е.  $\mathbf{E}(z) = \mathbf{T}_A$ , а  $\mathbf{R}(z) = \mathbf{T}_S$ . Например, восьмиканальный ( $M = 8$ ) максимально децимированный ПУБФ ассоциируется с другим преобразованием, широко используемым в трансформационном кодировании изображений в стандарте JPEG, дискретным косинусным преобразованием (ДКП):

$$[\mathbf{C}]_{m,n} = \sqrt{\frac{2}{M}} \cdot c_m \cos\left(\frac{m(n+\frac{1}{2})\pi}{M}\right), [\mathbf{D}]_{m,n} = \sqrt{\frac{2}{M}} \cdot c_n \cos\left(\frac{n(m+\frac{1}{2})\pi}{M}\right),$$

где  $\mathbf{D} = \mathbf{C}^{-1} = \mathbf{C}^T$ ;

$$c_m = \begin{cases} \frac{1}{\sqrt{2}} & (m = 0), \\ 1 & (m \neq 0), \end{cases} \quad c_n = \begin{cases} \frac{1}{\sqrt{2}} & (n = 0), \\ 1 & (n \neq 0); \end{cases}$$

$0 \leq m, n \leq M - 1$ ,  $m$  – количество столбцов,  $n$  – количество строк;

$[\mathbf{C}]_{m,n}$  – матрица прямого ДКП II второго типа;

$[\mathbf{D}]_{m,n}$  – матрица ДКП III третьего типа, являющаяся обратной для ДКП II второго типа (ОДКП).

Для простоты полагается, что  $M = 2^n$  ( $n \in \mathbb{N}$ ). Когда «cos» меняется на «sin», то преобразование называется дискретным синусным преобразованием (ДСП).



## 4.2 Необходимые свойства декоррелирующего преобразования в системе кодирования изображений

Как показано выше, банки фильтров можно рассматривать как обобщенный преобразователь в системе трансформационного кодирования изображений, которые должны иметь следующие необходимые свойства:

1) *Перфективную реконструкцию* исходного изображения (банк фильтров биортогональный (БОБФ) или ортогональный (ПУБФ)). Свойство перфективной реконструкции обеспечивает представление исходного изображения без потерь, высокий показатель эффективности кодирования и в случае ортогонального преобразователя преимущество заключается в том, что энергия ошибки в восстановленном изображении равна средней энергии ошибки в субполосах, что значительно упрощает анализ ошибок преобразования и позволяет использовать оптимальные алгоритмы распределения битов. Однако у большинства систем трансформационного кодирования изображений преобразователи биортогональные. Выполнение же ограничения ортогональности позволит обеспечить столь необходимую гибкость в проектировании преобразователя.

2) При обработке изображения крайне важно, чтобы у всех фильтров анализа и синтеза была *линейная ФЧХ*. Помимо устранения искажения фазы, линейность ФЧХ позволяет использовать простые симметричные методы расширения блоков анализа, чтобы точно обработать границы блоков изображений, а также осуществить более эффективную реализацию банка фильтров.

3) *Длина импульсной характеристики фильтра* прототипа должна быть относительно короткой, чтобы предотвратить «звонящие» артефакты в восстановленных изображениях и сохранить возможность синтеза быстрого преобразования.

4) Так как интерес представляет степень сжатия, т. е. получение низкой скорости битового потока, то предпочтительнее *максимально децимированный банк фильтров*.

5) Высокий уровень концентрации энергии в субполосах преобразователя. Показатель *эффективности кодирования CG* (coding gain) характеризует степень компактности энергии сигнала, прошедшего через банк фильтров, и определяется следующим образом:

$$CG = 10 \lg \left[ \frac{\sigma_x^2}{\left[ \prod_{k=0}^{M-1} \sigma_{xk}^2 \|\Phi_k\|^2 \right]^{1/M}} \right], \quad (4.3)$$

где  $\sigma_x^2$  – дисперсия входного сигнала;

$\sigma_{xk}^2$  – дисперсия сигнала в  $k$ -й субполосе банка анализа с базовой функцией преобразования  $\Phi_k$ ;

$\|\Phi_k\|^2$  – квадрат нормы базовой функции  $\Phi_k$   $k$ -й субполосы банка синтеза преобразователя.

В случае *ортогонального преобразователя*, ПУБФ, эффективность субполосного кодирования определяется отношением среднего арифметического к среднему геометрическому неотрицательных величин  $\sigma_{xk}^2$ :

$$CG = 10 \lg \left[ \frac{1/M \sum_{k=0}^{M-1} \sigma_{xk}^2}{\left[ \prod_{k=0}^{M-1} \sigma_{xk}^2 \right]^{1/M}} \right]. \quad (4.4)$$

Так как среднее арифметическое больше или равно среднему геометрическому, то эффективность субполосного кодирования  $CG \geq 1$ , т. е. субполосный кодер обеспечивает равное или лучшее качество по сравнению с широкополосным кодером при любом входном сигнале.

б) Хорошая частотная избирательность – максимальное ослабление в полосе непропускания банка фильтров, величина которой определяется соответственно как для банков анализа:

$$E_A = \sum_{k=0}^{M-1} \int_{\omega \in \Omega_k} |\Phi_k(e^{j\omega})|^2 d\omega \text{ и синтеза } E_S = \sum_{k=0}^{M-1} \int_{\omega \in \Omega_k} |\Phi_k(e^{j\omega})|^2 d\omega.$$

7) Сохранение структурной регулярности и отсутствие постоянной составляющей в субполосах преобразования кроме низкочастотной субполосы  $H_0(z)$ , т. е. эффекта DC leakage. Для БОБФ регулярность определяется парой порядков для анализа  $K_A$  и синтеза  $K_S$ , но для ПУБФ  $K_A = K_S \triangleq K$ . 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  $K$ -й порядок регулярности, полифазная матрица  $\mathbf{E}(z)$  должна удовлетворять условию  $\frac{d^n}{dz^n} \left\{ \mathbf{E}(z^M) \left[ 1 \ z^{-1} \ \dots \ z^{-(M-1)} \right]^T \right\}_{z=1} = c_n \mathbf{e}$ , где  $c_n \neq 0$  для  $n=0, \dots, K-1$ ,  $\mathbf{e} = [1 \ 0 \ 0 \ 0]^T$ . Например, в  $M$ -канальном банке фильтров с регулярностью 1-го порядка ( $K=1$ ) значения амплитудно-частотных характеристик всех канальных фильтров  $H_k(z)$ ,  $k=1, \dots, M-1$  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 обрабо-

танном изображении таким банком фильтров отсутствуют артефакты «шахматной доски» (checker board artifact), а также выше степень компрессии, потому что энергия в субполосах концентрируется компактно. На рисунке 4.3 показаны изображения, реконструированные ПУБФ без выполнения условия регулярности первого порядка (рисунок 4.3, а) и ПУБФ с регулярностью первого порядка (рисунок 4.3, б). На рисунке 4.3, а четко видна сетка поверх образа, представляющая собой артефакт «шахматной доски», а реконструированное изображение на рисунке 4.3, б является чистым.

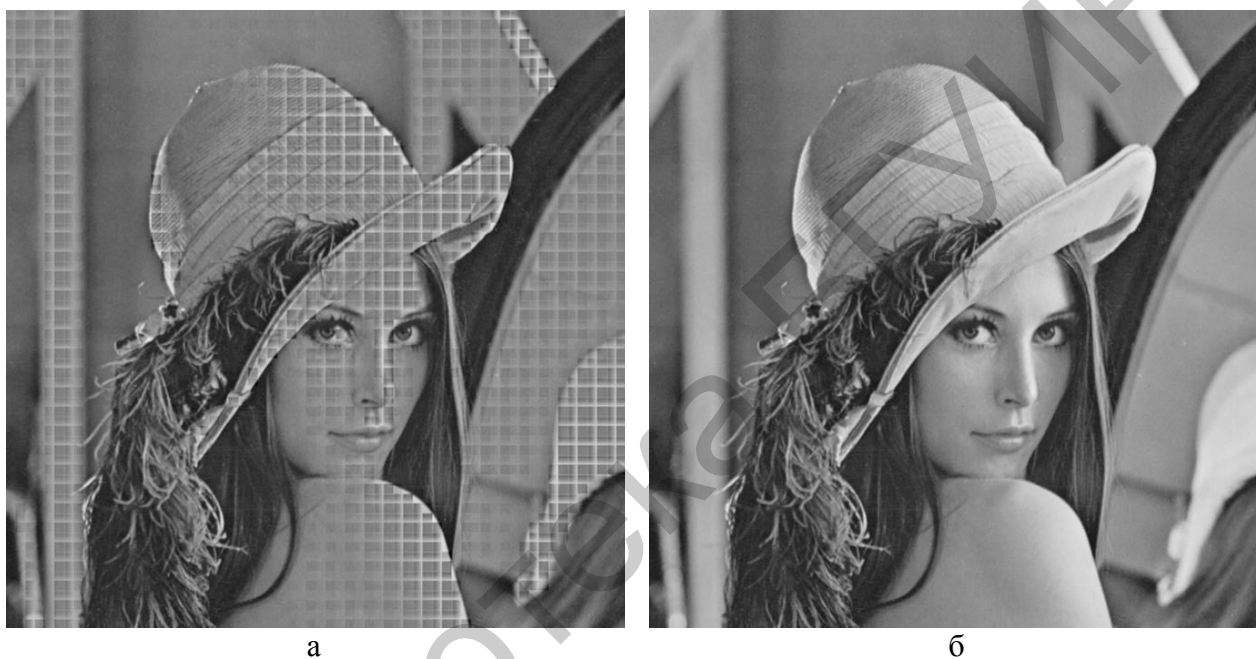


Рисунок 4.3 – Реконструированные изображения ПУБФ

8) Обеспечение заданной гладкости базовых функций  $\Phi_k$  и  $\Phi_k$  для хорошего перцептуального восприятия реконструированного изображения  $X$ .

9) Представление коэффициентов  $U$  преобразования  $T$  в формате фиксированной запятой, т. к. оцифрованные изображения имеют фиксированный формат пикселей, а также с целью упрощения аппаратной реализации.

10) Обеспечение обратимых «целое к целому» преобразований исходного изображения  $X$  и недопущение расширения динамического диапазона представления коэффициентов  $U$  преобразования  $T$ . Данные два свойства критически важны для приложений кодирования без потерь.

11) Структура преобразователя  $T$  должна достаточно просто отображаться на аппаратное и программное обеспечение: структурная регулярность, модульность, высокий вычислительный параллелизм, буферы памяти небольшой емкости, малая латентность и потребляемая мощность.

### 4.3 Мотивация построения кодера изображений по схеме L2L: lossless-to-lossy

Кодирование изображения (сигнала) без потерь (lossless) обычно относится к способам кодирования, которые полностью обратимы, т. е. кодеры, где исходное изображение (сигнал) может быть реконструировано бит к биту. Во многих приложениях компрессия без потерь есть единственно допустимый способ сокращения объема данных. Одним из таких приложений является архивация медицинских и деловых документов, сжатие с потерями которых обычно запрещено по закону. Другое приложение – это обработка спутниковых изображений, где как применение, так и стоимость получения исходных данных делают сжатие нежелательным. Еще одним направлением использования компрессии без потерь является цифровая рентгенология, в которой потеря информации может ухудшить точность диагностики. В этих и других областях необходимость применения сжатия изображений без потерь продиктована особенностями использования обработанных изображений.

Кодирование изображения (сигнала) с потерями (lossy) обычно относится к кодерам, которые создают приблизительное восстановление исходного изображения (сигнала). Характер потерь полностью зависит от используемого способа кодирования с потерями. При этом основным критерием эффективности выступает оптимальное соотношение степени компрессии и качества восстановления передаваемых или хранимых данных, в качестве которых часто выступают такие мультимедаданные, как изображения, видео, аудио и речевые сигналы. Это соотношение может зависеть от множества факторов, начиная от характеристик канала связи (пропускная способность) и заканчивая сферой применения систем обработки данных. Например, применение трансформационного кодирования изображений в бытовых системах мультимедиа: сотовая связь, IP-телефония, фотоаппараты, видеокамеры и допускает внесение определенного количества искажений с целью уменьшения битового потока (увеличения степени сжатия).

Особый интерес в настоящее время представляют схемы, работающие по принципу L2L (lossless-to-lossy) и осуществляющие сжатие и восстановление мультимедаданных как без потерь, так и с контролируемым внесением артефактов. Схема кодирования L2L унифицирует схемы кодирования изображений lossy и lossless, когда требуется высокое качество и степень компрессии. Это обеспечивает возможность широкого применения схем компрессии изображений в соответствии с мультимедийными приложениями, а задача создания универсального кодера изображений, работающего по схеме L2L (lossless-to-lossy), в настоящее время весьма актуальна.

Обычно при кодирование изображений по схеме с потерями (lossy) и трансформационное преобразование **T** построены на банках фильтров: ДКП в JPEG,

ортогональные преобразования с перекрытием LOT и GenLOT, 9/7 – tap ДВП в JPEG 2000, хорошо изучены. Данные банки фильтров реализованы на арифметике с плавающей запятой, и, соответственно, коэффициенты преобразования  $u$  представлены в формате плавающей запятой, в то время как в оцифрованном изображении пиксели упакованы в формат фиксированной запятой. Таким образом, коэффициенты  $u$  преобразования субполосного кодера квантуются по сценарию lossy (с потерями). Следовательно, восстановленное изображение далеко не соответствует оригиналу (исходному изображению) даже несмотря на то, что используется высокая скорость битового потока (низкая степень сжатия). Поэтому такой подход не может быть положен в основу компрессии изображений по схеме L2L (lossless-to-lossy).

Для кодирования изображения без потерь (схема lossless) требуется, чтобы трансформационное преобразование было обратимым и реализовано в целочисленной арифметике, т. е. преобразование соответствовало схеме обработки «целое в целое». С другой стороны, для применения кодеров в системах мультимедиа реального времени необходима аппроксимация коэффициентов преобразования с плавающей запятой в формат арифметики с фиксированной запятой. Целочисленные ДКП (IntDCT), которые реализованы на основе лестничной схемной параметризации, были предложены для трансформационного кодирования изображений. Однако они не могут быть использованы для кодирования изображений по схеме без потерь (lossless), потому что динамический диапазон представления коэффициентов лестничной схемной параметризации не контролируется и часто бывает очень большим, что требует большой разрядности слова представления коэффициентов или необходимо реализовывать арифметику с фиксированной запятой переменного формата. Решение также неприемлемо для кодирования изображения по схеме без потерь (lossless) при реализации кодера в арифметике с малой разрядностью из-за структуры представления коэффициентов.

Стандарты JPEG, JPEG2000 уже имеют режимы сжатия по схемам lossy и lossless: независимо осуществляется кодирование по схеме lossy (с потерями) и кодирование исходного изображения по схеме lossless (без потерь). Такой подход предполагает выполнение трансформационного преобразования дважды, что увеличивает вычислительную сложность кодера. Простейший путь к кодированию изображения по схеме L2L – это применить целочисленные преобразователи, например дискретное вейвлет-преобразование 5/3 – tap стандарта JPEG2000 или 4×8 – иерархическое расширенное биортогональное преобразование (HLBT) стандарта HDPhoto (JPEG-XR). Однако данные трансформационные преобразователи проектировались для режима сжатия исходного изображения

по схеме lossless и не соответствуют кодированию изображений по схеме lossy (с потерями).

В работах предложены структуры банков фильтров на основе лестничной схемной параметризации для кодирования изображения по принципу L2L, но в большинстве данных работ не берется во внимание число операций округления результатов умножения в лестничной структуре, которые можно рассматривать как белый шум квантования. Число операций округления влияет на компактность энергии в субполосах банка фильтров и получить высокую эффективность систем сжатия изображений по схеме L2L не представляется возможным. Следовательно, количество операций округления в целочисленном трансформационном преобразователе для схемы L2L компрессии изображений должно быть уменьшено в максимально возможной степени.

## 4.4 Структурный синтез преобразователей для систем трансформационного кодирования изображений

### 4.4.1 Блочные лестничные структурные преобразования

С позиций банка фильтров вейвлет-преобразование и модифицированное ДКП (расширенное) просто представляются как системы высокого порядка, у которых полифазные матрицы состоят из полиномов переменной  $z$ . В то время как блочные преобразования на основе ДКП и целочисленного варианта ДКП описываются матрицами со скалярными компонентами, т. е. полиномами нулевого порядка. В случае вейвлет-преобразования это двухканальный банк фильтров, и в общем случае проектирование базируется на факторизации фильтра нижних частот. Однако когда число каналов  $M$  увеличивается, то спектральная факторизация (построение диадического дерева вейвлет-преобразования) не всегда практична.

Наиболее успешный подход к построению многоканального банка фильтров или многополосного вейвлет-преобразования основан на *решетчатой схемной параметризации* полифазной матрицы  $\mathbf{E}(z)$ . Здесь полифазная матрица  $\mathbf{E}(z)$   $M$ -канального банка фильтров анализа с совершенной реконструкцией, компоненты которой полиномы высокого порядка, представляется как каскадная композиция множества блоков меньшего порядка  $\mathbf{E}_0$  и  $\mathbf{R}_i$ , каждый из которых удовлетворяет необходимым свойствам преобразователя в системе кодирования изображений, таким, как симметрия импульсной характеристики, совершенная реконструкция и отсутствие генерации постоянной составляющей (DC leakage):

$$\mathbf{E}(z) = \mathbf{R}_{N-1}\Lambda(z)\mathbf{R}_{N-2}\Lambda(z)\dots\mathbf{R}_1\Lambda(z)\mathbf{E}_0 = \left(\prod_{i=N-1}^1 \mathbf{R}_i\Lambda(z)\right)\mathbf{E}_0, \quad (4.5)$$

где  $MN (N \in \mathbb{N})$  – длина импульсной характеристики фильтра прототипа;

$\Lambda(z) = \text{diag}(z^{-1}, \mathbf{I}_{M-1})$  – матрица задержек, произвольные матрицы: общая  $\mathbf{E}_0$ , с  $M(M-1)/2$  степенями свободы и несингулярные матрицы ( $\mathbf{R}_i, i = 1 \dots N-1$ ), ограниченные до  $M-1$  степени свободы.

На рисунке 4.4 показана решетчатая структура банка фильтров анализа с совершенной реконструкцией.

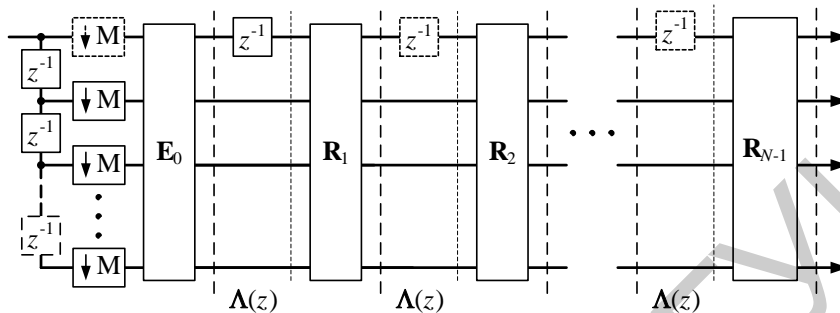


Рисунок 4.4 – Решетчатая структура банка фильтров анализа с совершенной реконструкцией

Когда  $\mathbf{R}_i$  – ортогональная матрица  $\mathbf{R}_i^{-1} = \mathbf{R}_i^T$ , тогда факторизация (4.5) соответствует ПУБФ. Например, оба типа матриц могут быть параметризованы планарными вращениями Гивенса, в которых каждый из них соответствует одной степени свободы.

Структурное проектирование системы банка фильтров направлено на поиск структуры, которая удовлетворяет как можно большему числу заявленных свойств банка фильтров. Другими словами, осуществляется оптимизация структуры блоков меньшего порядка  $\mathbf{R}_i$  в ограничениях целочисленной арифметики с целью получить максимальные оценки эффективности субполосного кодирования и ослабления в полосе непропускания банка фильтров.

В последние несколько десятилетий *лестничная схемная параметризация* дает возможность отображать банки фильтров (например, дискретное вейвлет преобразование) на пространство целых чисел и получать эффективные структуры преобразователей с совершенной реконструкцией. Следовательно, можно реализовать преобразователь, работающий по схеме L2L. Добыши и Свилденсон доказали, что любой фильтр с конечной импульсной характеристикой может быть факторизован в произведение конечного числа простых ниже- и выше-треугольных матриц. На рисунке 4.5 показана структура анализа-синтеза лестничной схемной параметризации.

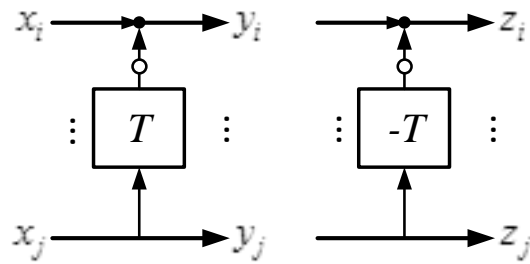


Рисунок 4.5 – Структура анализа-синтеза лестничной схемной параметризации (черные и белые кружки обозначают операции суммирования и округления соответственно)

Сигналы  $x_i$  и  $x_j$  на входе и на выходе  $y_i$  и  $y_j$  анализирующей части, а также сигналы  $z_i$  и  $z_j$  на выходе синтезирующей части и коэффициент  $T$  лестничной схемной параметризации соотносятся следующим образом:

$$\begin{aligned} y_i &= x_i + \text{round}[Tx_j] \rightarrow z_i = y_i - \text{raund}[Ty_j], \\ y_j &= x_j, \dots, z_j = y_j = x_j, \end{aligned} \quad (4.6)$$

где  $\text{raund}[\bullet]$  – оператор округления.

В общем случае  $T$  – это оператор, который может принимать различные значения в зависимости от того, насколько должна быть идеальна реконструкция. Он может быть любым скалярным числом, многочленом с задержкой, а также любым нелинейным оператором. В данном случае матрицы лестничной схемной параметризации и их инверсные матрицы имеют следующий вид:

$$\begin{bmatrix} 1 & T_0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & T_0 \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -T_0 \\ 0 & 1 \end{bmatrix}. \quad (4.7)$$

Это полностью обратимые структуры, поэтому, чтобы инвертировать шаг лестничной схемной параметризации, необходимо вычесть (или добавить) то, что было добавлено (или вычтено) в прямом преобразовании (анализирующей части). Одним из преимуществ предлагаемых структур является универсальность и простота построения быстрых преобразований. Однако здесь не берется во внимание число операций округления результатов умножения, рассматриваемых как белый шум, число которых влияет на компактность энергии в субполосах банка фильтров.

Предлагается новый класс *блочных лестничных структурных преобразований* для банка фильтров в системе трансформационного кодирования изображений по схеме L2L, на которых можно строить преобразователи, позволяющие получать коэффициент компрессии выше, чем на основе стандартных базовых лестничных структур. Техника блочных лестничных структурных преобразований позволяет уменьшить число узлов округления путем объединения многих



операций округления. На рисунке 4.6, а показана структура анализа-синтеза системы банка фильтров на основе лестничной схемной параметризации, а на рисунке 4.6, б – структура анализа-синтеза системы банка фильтров на базе блочных лестничных структурных преобразований.

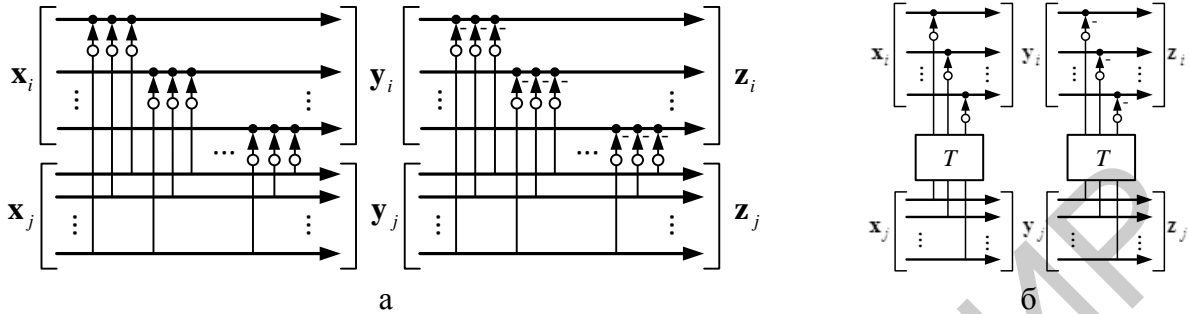


Рисунок 4.6 – Структура анализа-синтеза банка фильтров на основе лестничной схемной параметризации (а) и блочных лестничных структурных преобразований (б) (черные и белые кружки обозначают операции суммирования и округления соответственно)

Векторы  $\mathbf{x}_i$  и  $\mathbf{x}_j$  на входе и на выходе  $\mathbf{y}_i$  и  $\mathbf{y}_j$  анализирующей части, а также векторы  $\mathbf{z}_i$  и  $\mathbf{z}_j$  на выходе синтезирующей части и блоки преобразования – матрицы  $\mathbf{T}$  и  $-\mathbf{T}$  соотносятся следующим образом:

$$\begin{aligned} y_i &= x_i + \text{round}[\mathbf{T}x_j] \rightarrow z_i = y_i - \text{raund}[\mathbf{T}y_j], \\ y_j &= x_j \dots \dots \dots z_j = y_j = x_j, \end{aligned} \quad (4.8)$$

где  $\text{raund}[\cdot]$  – оператор округления.

Операции округления производятся не после каждой лестничной ступени, а после блоков преобразований, т. е. после умножения на матрицу преобразований  $\mathbf{T}$ . При этом число операций округления уменьшается с  $M^2$  до  $M$ , где размер матрицы коэффициентов блока лестничного преобразования  $\mathbf{T}$  равен  $M \times M$ .

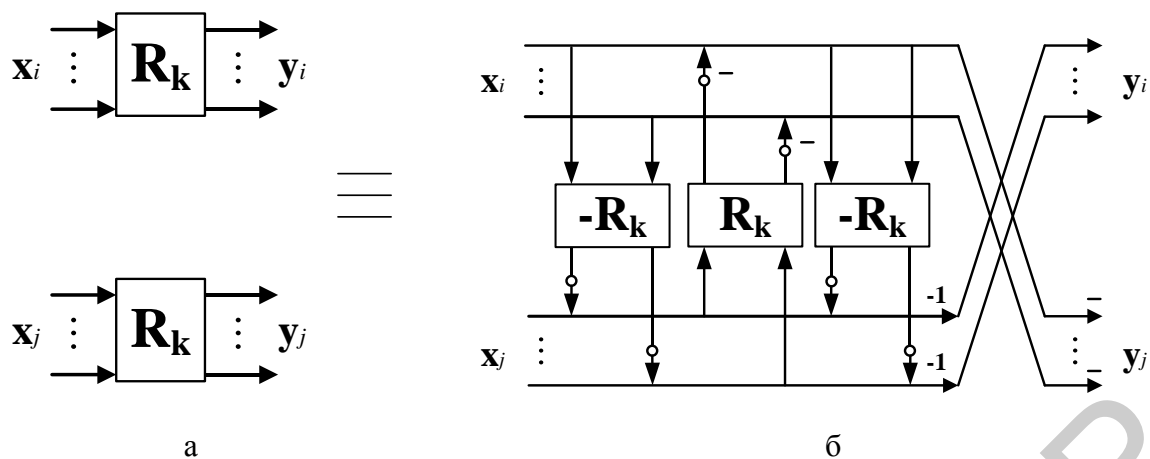
Как было показано выше, решетчатая структура банка фильтров может отражать пространственно-временной процесс обработки входного сигнала. Обычно используются *параллельно-блочные лестничные структурные преобразования* в системе банка фильтров, при этом два блока обрабатываются как один:

$$\mathbf{T}_k = \begin{bmatrix} \mathbf{R}_k & 0 \\ 0 & \mathbf{R}_k \end{bmatrix}. \quad (4.9)$$

Параллельно-блочное лестничное структурное преобразование выражения (4.9) представляется следующим образом:

$$\mathbf{T}_k = \begin{bmatrix} 0 & -\mathbf{I} \\ \mathbf{I} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \\ -\mathbf{R}_k & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{R}_k \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \\ -\mathbf{R}_k & \mathbf{I} \end{bmatrix}. \quad (4.10)$$

Блочная параллельная система  $\mathbf{T}_k$  и ее параллельно-блочное лестничное структурное преобразование показано на рисунке 4.7, а и б соответственно.



а – базовая структура; б – на основе параллельно-блочного лестничного структурного преобразования реконструкцией

Рисунок 4.7 – Блочно-параллельная структура банка фильтров с совершенной реконструкцией

Из выражения (4.10) видно, что в результате блочных лестничных структурных преобразований каждый блок  $T_k$  состоит из трех блоков лестничной структуры, представляющих собой матрицы коэффициентов лестничной схемной параметризации. Как видно из рисунка 4.7, много операций округления объединены вследствие матричных операций в структуре банка фильтров с совершенной реконструкцией. Таким образом, параллельно-блочное лестничное структурное преобразование (4.10) позволяет строить эффективные параллельные системы (4.9) без потерь, которые можно будет использовать в кодировании изображения по схеме L2L. Следует отметить, что в качестве операторов  $R_k$  и  $-R_k$  в данной структуре могут выступать матрицы любых преобразований, таких, как быстрое преобразование Фурье (БПФ), Уолша – Адамара, дискретное косинусное преобразование (ДКП).

#### 4.4.2 Решетчатые структуры параунитарных банков фильтров (ПУБФ) с линейными фазочастотными характеристиками

Для сохранения полезных свойств трансформационного преобразования важной является задача проектирования  $M$ -канальных ПУБФ с *линейными фазочастотными характеристиками* на основе решетчатой схемной параметризации. Для ПУБФ-анализа с линейной ФЧХ и четного числа каналов  $M$  в факторизации (4.5) передаточной полифазной матрицы  $E(z)$  выбор параметров компонент разложения осуществляется следующим образом:

$$E(z) = G_{N-1}(z)G_{N-2}(z)\dots G_1(z)E_0 = \left(\prod_{i=N-1}^1 G_i(z)\right)E_0, \quad (4.11)$$

где  $\mathbf{E}_0 = \frac{1}{\sqrt{2}} \Phi_0 \mathbf{W} \text{diag}(\mathbf{I}_{M/2}, \mathbf{J}_{M/2})$ ,  $\mathbf{G}_i(z) = \frac{1}{2} \Phi_i \mathbf{W} \Lambda(z) \mathbf{W}$ ,  $i=1, \dots, N-1$ ;

$$\mathbf{W} = \begin{bmatrix} \mathbf{I}_{M/2} & \mathbf{I}_{M/2} \\ \mathbf{I}_{M/2} & -\mathbf{I}_{M/2} \end{bmatrix}, \quad \Lambda(z) = \text{diag}(\mathbf{I}_{M/2}, z^{-1} \mathbf{I}_{M/2}), \quad \Phi_i = \text{diag}(\mathbf{U}_i, \mathbf{V}_i);$$

$\mathbf{U}_i$  и  $\mathbf{V}_i$  – произвольные ортогональные матрицы размером  $M/2 \times M/2$ .

Кроме того, для упрощения вышеобозначенной факторизации для  $i > 0$  матрица  $\mathbf{U}_i$  может быть представлена единичной матрицей, тогда

$$\Phi_i = \text{diag}(\mathbf{I}_{M/2}, \mathbf{V}_i), \quad i > 0. \quad (4.12)$$

Решетчатая структура четырехканального банка фильтров анализа с линейной ФЧХ изображена на рисунке 4.8 для  $N = 2$ .

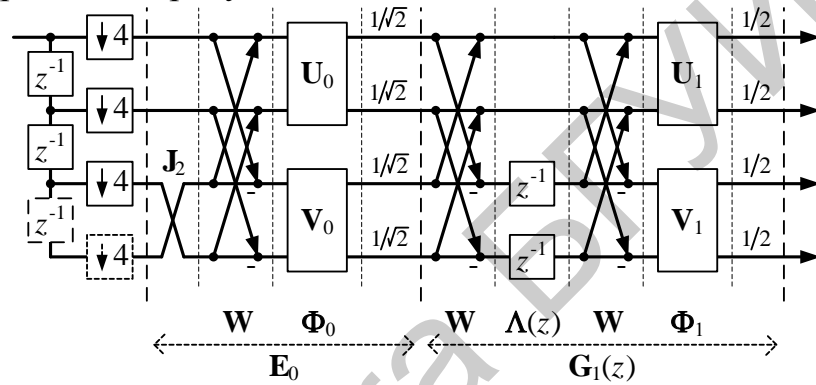


Рисунок 4.8 – Решетчатая структура четырехканального ПУБФ-анализа с линейной ФЧХ

Полифазная передаточная матрица  $\mathbf{E}(z)$  будет иметь свойство совершенной реконструкции, если все ее элементы  $\mathbf{G}_i(z)$  в факторизации (4.11) остаются унитарными (ортогональными) после квантования их коэффициентов. Свойство унитарности матрицы  $\mathbf{E}(z)$  эквивалентно тому, что строки  $\mathbf{r}_k$ ,  $k=0, \dots, M-1$  должны удовлетворять условию

$$\mathbf{r}_m^T \mathbf{r}_n = 0, \quad m \neq n \quad (\text{унитарность}),$$

$$\mathbf{r}_m^T \mathbf{r}_m = d > 0 \quad (\text{норма-константа}).$$

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

$$H_{M-1-k}(z) = \pm H_k(-z) \quad \text{или} \quad h_{M-1-k}(n) = \pm (-1)^n h_k(n), \quad (4.13)$$

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 получен, если модифицировать факторизацию решетчатой структуры ПУБФ с линейной ФЧХ (4.11). Действительно, достаточно ассоциировать  $U_i$  с  $V_i$  в (4.11) и представить матрицу  $U_i$  следующим образом:

$$\begin{aligned} U_i &= \Gamma V_i \Gamma, \quad i = 0, \dots, N-2, \\ U_{N-1} &= \mathbf{J}_{M/2} V_{N-1} \Gamma, \end{aligned} \quad (4.14)$$

где  $\mathbf{J}_{M/2}$  – обратная единичная матрица;

$\Gamma$  – диагональная матрица, элементы которой определяются как  $\gamma_{mm} = (-1)^{m-1}$ ,  $m = 1, \dots, M/2$ .

Количество степеней свободы при синтезе данного банка фильтров сократилось. Как видно из (4.14), параметры настройки банка фильтров определяются только матрицей  $V_i$ . Следовательно, оптимизация коэффициентов упростилась, что явилось основной мотивацией разработки ПУБФ с линейной ФЧХ, у которых наблюдается попарная зеркальная симметрия частотных характеристик фильтров относительно  $\pi/2$ .

Структура банка синтеза ПУБФ конструируется на основе обратного включения компонент факторизации полифазной передаточной матрицы банка анализа.

При синтезе и реализации ПУБФ с линейной ФЧХ матрицы  $U_i$  и  $V_i$  в разложении (4.11) и (4.12), как правило, параметризуются планарными оборотами Гивенса, каждый из которых соответствует одной степени свободы (рисунок 4.9). Данный подход подобен QR-декомпозиции матриц и приводит к нескольким эквивалентным вычислительным схемам. Например, для матрицы размером  $4 \times 4$  ( $M = 4$ ) на рисунке 4.9 показаны две типичных структуры параметризации планарными оборотами Гивенса. Углы вращения в матрице Гивенса могут быть выбраны таким образом, чтобы максимизировать оценку эффективности субполосного кодирования. Благодаря периодичности тригонометрических функций, это можно сделать с помощью эффективных алгоритмов оптимизации без ограничений.

Однако описанные выше факторизации теряют свойство параунитарности, когда матрицы  $U_i$  и  $V_i$  представляются с ограниченной точностью. Перфективная реконструкция не обеспечивается банком фильтров и решетчатыми структурами с фиксированной запятой, потому что наличия в разложении (4.11) даже одной неортогональной компоненты достаточно, чтобы условие перфективной реконструкции не выполнялось. Это обусловлено тем, что квантование компонент ортогональных матриц в общем случае приведет к изменению норм столбцов.

$$\begin{bmatrix} 1 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 \\ 0 & \cdots & 0 & \cos \alpha & -\sin \alpha \\ 0 & \cdots & 0 & \sin \alpha & \cos \alpha \end{bmatrix}$$

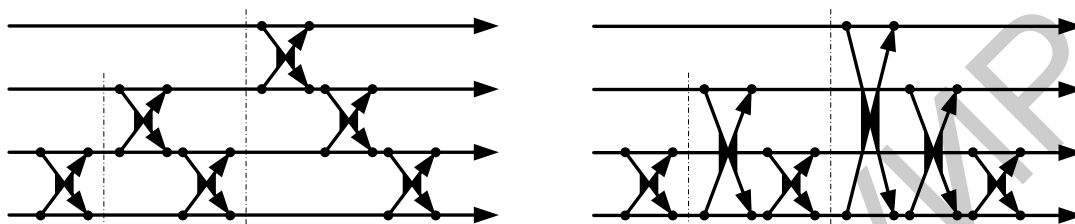
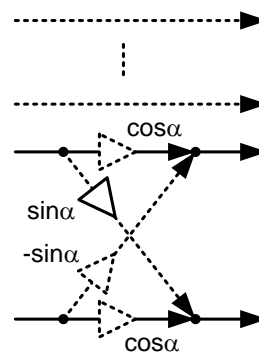


Рисунок 4.9 – Структуры параметризации планарными оборотами Гивенса

Например, квантованная матрица оборотов Гивенса

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & Q(\cos \alpha) & -Q(\sin \alpha) & 0 \\ 0 & Q(\sin \alpha) & Q(\cos \alpha) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (4.15)$$

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

Анализ решетчатой параметризации блочно-диагональной матрицы  $\Phi_i$  показывает, что выражение условия регулярности первого порядка может быть включено в процедуру синтеза ПУБФ. Это может быть сделано в терминах углов вращения решетчатой параметризации полифазной матрицы  $E(z)$ . Однако данное решение не учитывает, что квантование матриц вращения Гивенса вызывает определенные отклонения углов от их первоначальных значений. Следует заметить, что диадическая параметризация (на основе отражений Хаусхолдера) и лестничная схемная параметризация также имеют данный недостаток, хотя последняя, как было показано ранее, гарантирует перфективную реконструкцию входного сигнала, но банки фильтров на основе лестничной схемной параметризации являются структурами БОБФ.

Таким образом, актуальной является задача поиска альтернативного способа проектирования ПУБФ с линейной ФЧХ как обратимого преобразователя в системах трансформационного кодирования изображений существующим классическим методам. Одним из таких решений преобразователя могут быть пара-

унитарные банки фильтров (ПУБФ) с линейной ФЧХ на основе алгебры кватернионов, которые по сути являются преобразованиями без потерь. Для числа каналов больше двух линейность ФЧХ соблюдается, лестничные и решетчатые структуры также сохраняют свойство совершенной реконструкции входного сигнала, что дает возможность использовать арифметику с фиксированной запятой.

## 4.5 Решетчатые структуры регулярного ПУБФ на основе алгебры кватернионов

### 4.5.1 Кватернионы и ортогональные матрицы

Алгебра кватернионов  $\mathbb{H}$  является ассоциативной некоммутативной четырехмерной алгеброй  $\mathbb{H} = \{\mathbf{q} = q_1 + q_2i + q_3j + q_4k \mid q_1, q_2, q_3, q_4 \in \mathbb{R}\}$ , где ортогональные мнимые части подчиняются следующим законам умножения:  $i^2 = j^2 = k^2 = ijk = -1$ ,  $ij = -ji = k$ ,  $jk = -kj = i$ ,  $ki = -ik = j$ . Операции сложения и умножения кватернионов могут быть реализованы с использованием векторно-матричной алгебры. Для этого кватернионы представляются в виде четырехмерных векторов. Умножение кватернионов в векторной форме производится по правилу умножения вектора на матрицу, так как операция умножения некоммутативна, то продукт умножения определяется операторами умножения «слева»  $\mathbf{M}^+(q)$  и «справа»  $\mathbf{M}^-(q)$ :

$$qx \Leftrightarrow \underbrace{\begin{bmatrix} q_1 & -q_2 & -q_3 & -q_4 \\ q_2 & q_1 & -q_4 & q_3 \\ q_3 & q_4 & q_1 & -q_2 \\ q_4 & -q_3 & q_2 & q_1 \end{bmatrix}}_{\mathbf{M}^+(q)} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_{\mathbf{x}}, \quad xq \Leftrightarrow \underbrace{\begin{bmatrix} q_1 & -q_2 & -q_3 & -q_4 \\ q_2 & q_1 & q_4 & -q_3 \\ q_3 & -q_4 & q_1 & q_2 \\ q_4 & q_3 & -q_2 & q_1 \end{bmatrix}}_{\mathbf{M}^-(q)} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}}_{\mathbf{x}}. \quad (4.16)$$

Обе матрицы (4.16) ортогональные до масштабного коэффициента, заданного нормой кватернионов  $|q| = \sqrt{q_1^2 + q_2^2 + q_3^2 + q_4^2}$ , и имеют детерминант, равный +1, что позволяет отнести их к группе специальных матриц размером  $4 \times 4$ , называемых  $SO(4)$ :

$$\mathbf{M}^\pm(q)^{-1} = (1/|q|)\mathbf{M}^\pm(q)^T,$$

где  $T$  – знак транспонирования матрицы.

Как следует из структур матриц, их свойства не зависят от значений компонент кватерниона и, следовательно, нечувствительны к их модификации, например квантованию. Матрицы  $\mathbf{M}^+(q)$  и  $\mathbf{M}^-(q)$  соотносятся между собой как

$$\mathbf{M}^\mp(q) = \mathbf{D}_C \mathbf{M}^\pm(q)^T \mathbf{D}_C,$$

где  $\mathbf{D}_C = \text{diag}(1, -\mathbf{I}_3)$  описывает в матричной нотации оператор гиперкомплексного сопряжения. Определив сопряженный кватернион  $\bar{q} = q_1 - q_2i - q_3j - q_4k$  в векторно-матричном виде как  $\bar{\mathbf{q}} = \mathbf{D}_C \mathbf{q}$ , можно получить, что

$$\mathbf{M}^{\pm}(\bar{q}) = \mathbf{M}^{\pm}(q)^T,$$

тогда

$$\mathbf{M}^{\pm}(\bar{q}) = \mathbf{D}_c \mathbf{M}^{\pm}(q) \mathbf{D}_c.$$

Последнее равенство показывает, что результаты, полученные для умножения «левого»  $qx$  или «правого» типа  $xq$ , могут быть применимы к умножению на сопряженный кватернион. В частности, это доказывает, что вычислительная сложность операторов умножений кватернионов одинакова. Для изображения операторов умножения  $\mathbf{M}^{-}(q)$  и  $\mathbf{M}^{+}(q)$  на структурных схемах могут быть использованы следующие графические обозначения (рисунок 4.10).

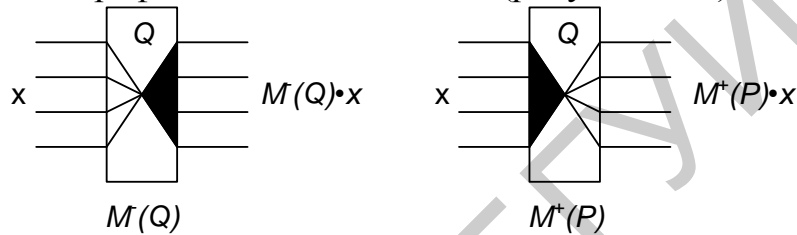


Рисунок 4.10 – Графические изображения операторов умножения кватернионов

Для матриц операторов умножения кватернионов верны следующие равенства:

$$\mathbf{M}^{+}(Q_{N-1})\mathbf{M}^{+}(Q_{N-2})\dots\mathbf{M}^{+}(Q_0) = \mathbf{M}^{+}(Q_{N-1}Q_{N-2}\dots Q_0), \quad (4.17)$$

$$\mathbf{M}^{-}(Q_{N-1})\mathbf{M}^{-}(Q_{N-2})\dots\mathbf{M}^{-}(Q_0) = \mathbf{M}^{-}(Q_0Q_1\dots Q_{N-1}). \quad (4.18)$$

Доказан ряд теорем для обоснования применения умножителей кватернионов в качестве составляющих блоков ПУБФ:

1) Для каждой ортогональной матрицы  $\mathbf{R}$  размером  $4 \times 4$  существует уникальная (с точностью до знака) пара единичных кватернионов  $P$  и  $Q$ , таких, что:

$$\forall_{\mathbf{R} \in SO(4)} \exists_{P, Q \in \text{unit quat}} \mathbf{R} = \mathbf{M}^{+}(P) \cdot \mathbf{M}^{-}(Q) = \mathbf{M}^{-}(Q) \cdot \mathbf{M}^{+}(P). \quad (4.19)$$

2) Блочная диагональная матрица размером  $8 \times 8$ , образованная из двух произвольных ортогональных матриц  $\mathbf{U}$  и  $\mathbf{V}$  размером  $4 \times 4$ , может быть описана следующим образом:

$$\begin{aligned} \text{diag}(\mathbf{U}, \mathbf{V}) &= \text{diag}(\mathbf{M}^{-}(S), \mathbf{M}^{-}(S)) \cdot \text{diag}(\mathbf{M}^{-}(\bar{R}), \mathbf{M}^{-}(R)) \times \\ &\times \text{diag}(\mathbf{M}^{+}(\bar{Q}), \mathbf{M}^{+}(Q)) \cdot \text{diag}(\mathbf{M}^{+}(P), \mathbf{M}^{+}(P)), \end{aligned} \quad (4.20)$$

где каждый элемент связан с одним из четырех единичных кватернионов  $P, Q, R, S$ . Когда оба блока диагональной матрицы равны ( $\mathbf{U} = \mathbf{V}$ ), выражение (4.20) примет следующий вид:

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

Если блок  $\mathbf{U}$  диагональной матрицы является единичной матрицей  $\mathbf{I}_4$ , то выражение (4.20) преобразуется к виду

$$\begin{aligned} \text{diag}(\mathbf{I}_4, \mathbf{V}) = & \text{diag}(\mathbf{M}^-(Q), \mathbf{M}^-(Q)) \cdot \text{diag}(\mathbf{M}^-(\bar{Q}), \mathbf{M}^-(Q)) \times \\ & \times \text{diag}(\mathbf{M}^+(\bar{P}), \mathbf{M}^+(P)) \cdot \text{diag}(\mathbf{M}^+(P), \mathbf{M}^+(P)). \end{aligned} \quad (4.21)$$

Используя графические изображения операторов умножения (рисунки 4.11), можно представить выражение (4.20) в виде следующей схемы.

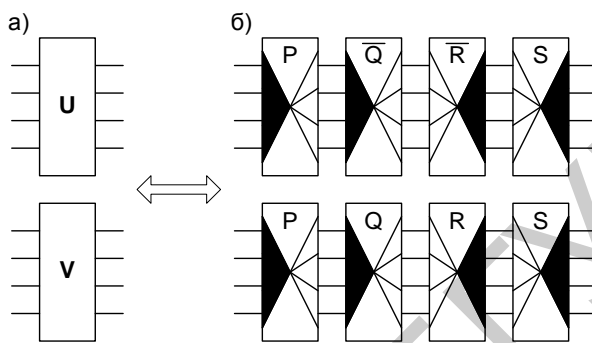


Рисунок 4.11 – Структурное изображение блочной диагональной матрицы размером  $8 \times 8$  (а) и реализация выражения (4.20) на умножителях кватернионов (б)

Факторизация (4.19) всегда соответствует 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:

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

Здесь  $Q$  – оператор квантования.

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 необязательно единичны, невзирая на конечную точность представления матриц умножения кватернионов  $\mathbf{M}^-(q)$  и  $\mathbf{M}^+(q)$ .

Таким образом, матрица  $\mathbf{R}$  в (4.19) всегда представляет ортогональное преобразование, а умножение кватернионов может рассматриваться как основной вычислительный блок ПУБФ. Произвольная матрица  $4 \times 4$  требует хранения 16 действительных чисел и 32 чисел для блочно-диагональной матрицы  $8 \times 8$ . В данном же случае необходимо запомнить в памяти 8 и 16 действительных чисел (2 и 4 кватерниона) соответственно. Последние двадцать лет в цифровой обработке сигналов алгебра кватернионов рассматривается как новая парадигма, позволяющая осуществлять обработку сигналов непосредственно в многомерном



доме. Кватернионы уже с успехом использовались в цифровой обработке сигналов, например, в адаптивных фильтрах Калмана, оценке спектра по методу MUSIC, сингулярной декомпозиции матриц, параунитарных банках фильтров (ПУБФ) – Q-ПУБФ и других приложениях.

#### 4.5.2 Решетчатая схемная параметризация ПУБФ на основе алгебры кватернионов

На основании утверждения (4.19) факторизация полифазных передаточных матриц анализа  $\mathbf{E}(z)$  и синтеза  $\mathbf{D}(z)$  решетчатой структуры четырехканального ПУБФ в алгебре кватернионов будет соответственно равна

$$\mathbf{E}(z) = \left( \prod_{i=N-1}^1 \mathbf{M}^\pm(Q_i) \Lambda(z) \right) \mathbf{M}^+(Q_0) \mathbf{M}^-(P_0), \quad (4.22)$$

$$\mathbf{D}(z) = \mathbf{M}^-(\bar{P}_0) \mathbf{M}^+(\bar{Q}_0) \left( \prod_{i=1}^{N-1} \Lambda(z) \mathbf{M}^\pm(\bar{Q}_i) \right), \quad (4.23)$$

где  $\bar{P}_0$  и  $\bar{Q}_0$  – единичные сопряженные кватернионы;

$\Lambda(z) = \text{diag}(z^{-1}, \mathbf{I}_{M-1})$  и  $\Lambda(z) = \text{diag}(1, z^{-1} \mathbf{I}_{M-1})$  – матрицы задержек;

$\mathbf{I}_{M-1}$  – обозначает единичную матрицу  $M \times M$ .

Решетчатые структуры анализа и синтеза четырехканального общего ПУБФ в алгебре кватернионов (Q-ПУБФ) для  $N = 3$  показаны на рисунке 4.12, а и б соответственно.

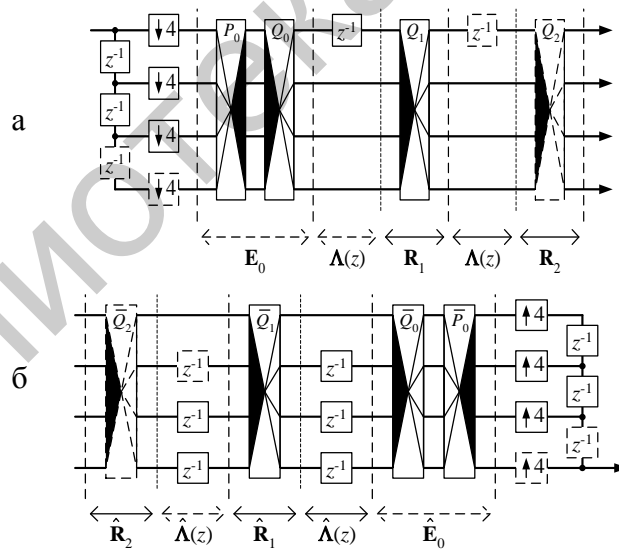


Рисунок 4.12 – Решетчатая структура четырехканального Q-ПУБФ анализа (а) и синтеза (б)

Элементы решетчатой факторизации критически децимированного четырехканального ПУБФ с линейной фазочастотной характеристикой (4.11) в алгебре кватернионов задаются следующим образом:

$$\Phi_0 = \mathbf{M}^-(P_0) \mathbf{M}^+(Q_0); \Phi_i = \mathbf{M}^-(P_i), \text{ при } i = 1, \dots, N, \quad (4.24)$$

где  $P_i, Q_0$  – некоторые единичные кватернионы.

Q-ПУБФ с линейной ФЧХ удовлетворяет условию регулярности 1-го порядка, если  $Q_0 = \pm \frac{1}{\sqrt{2}} \overline{P_0 \cdot P_1 \cdot \dots \cdot P_N} \cdot A$ , где  $A$  – специальный кватернион, имеющий реальную и одну мнимую часть  $A = 1 + i$ . На основании (4.11) и (4.24) решетчатая схемная параметризация передаточных матриц анализа  $\mathbf{E}(z)$  и синтеза  $\mathbf{D}(z)$  Q-ПУБФ соответственно определяются как

$$\mathbf{E}(z) = \left( \prod_{i=N-1}^1 \frac{1}{2} \mathbf{M}^-(P_i) \mathbf{W} \Lambda(z) \mathbf{W} \right) \frac{1}{\sqrt{2}} \mathbf{M}^-(P_0) \mathbf{M}^+(Q_0), \quad (4.25)$$

$$\mathbf{D}(z) = \frac{1}{\sqrt{2}} \mathbf{M}^-(\overline{P_0}) \mathbf{M}^+(\overline{Q_0}) \left( \prod_{i=1}^{N-1} \frac{1}{2} \mathbf{M}^-(\overline{P_i}) \mathbf{W} \Lambda(z) \mathbf{W} \right), \quad (4.26)$$

где  $\overline{P_0}$  и  $\overline{Q_i}$  – сопряженные единичные кватернионы;

$$\mathbf{W} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{I}_2 \\ \mathbf{I}_2 & -\mathbf{I}_2 \end{bmatrix}, \Lambda(z) = \text{diag}(\mathbf{I}_2, z^{-1} \mathbf{I}_2);$$

$$\Lambda(z) = \text{diag}(z^{-1} \mathbf{I}_2, \mathbf{I}_2).$$

Структура разложения Q-ПУБФ (4.25) для  $N = 3$  приведена на рисунке 4.13.

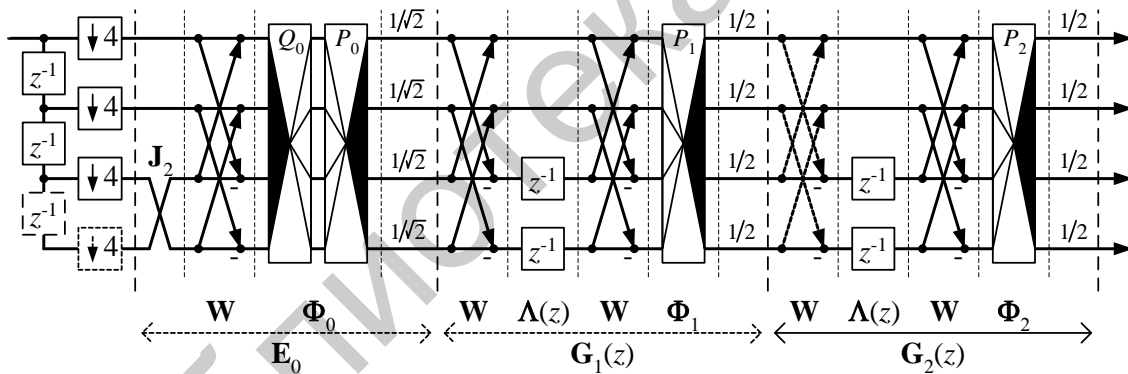


Рисунок 4.13 – Решетчатая структура анализа четырехканального Q-ПУБФ с линейной ФЧХ

Матрицы в разложении (4.11) для анализа Q-ПУБФ с попарной зеркальной симметрией частотных характеристик фильтров относительно  $\pi/2$  должны задаваться как

$$\Phi_N = \mathbf{M}^-(P_N) \text{diag}(\mathbf{J}_2 \Gamma, \mathbf{I}_2); \Phi_i = \mathbf{M}^-(P_i), \text{ при } i = 0, \dots, N-1, \quad (4.27)$$

где  $P_i$  – некоторые единичные кватернионы, которые должны удовлетворять следующему равенству (регулярность 1-го порядка):

$$P_N = \pm \frac{1}{\sqrt{2}} A \cdot \overline{P_0 \cdot P_1 \cdot \dots \cdot P_{N-1}}.$$

На основании утверждений (4.20) – (4.21) элементы решетчатой схемной параметризации в (4.11) для системы анализа восьмиканального Q-ПУБФ с линейной ФЧХ задаются следующим образом:

$$\Phi_0 = \text{diag}(\mathbf{M}^-(S_0), \mathbf{M}^-(S_0)) \cdot \text{diag}(\mathbf{M}^-(\bar{R}_0), \mathbf{M}^-(R_0)) \times \\ \times \text{diag}(\mathbf{M}^+(\bar{Q}_0), \mathbf{M}^+(Q_0)) \cdot \text{diag}(\mathbf{M}^+(P_0), \mathbf{M}^+(P_0)), \quad (4.28)$$

$$\Phi_i = \text{diag}(\mathbf{M}^-(\bar{R}_i), \mathbf{M}^-(R_i)) \cdot \text{diag}(\mathbf{M}^+(\bar{Q}_i), \mathbf{M}^+(Q_i)), \text{ при } i = 1, \dots, N, \quad (4.29)$$

где  $S_0, Q_i, R_i, P_0$  – некоторые единичные кватернионы, связанные следующим соотношением  $R_N = \frac{1}{2} \bar{Q}_N \cdot \bar{Q}_{N-1} \dots \bar{Q}_0 \cdot P_0 \cdot \bar{R}_0 \cdot S_0 \cdot \bar{R}_1 \dots \bar{R}_{N-1}$ , для выполнения условия регулярности 1-го порядка.

Таким образом, в решетчатых структурах Q-ПУБФ условие регулярности первого порядка может быть непосредственно зафиксировано в значениях коэффициентов при синтезе банка фильтров, например, в отличие от ПУБФ на основе ДКП – ОДКП, где используется дополнительный информационный блок размером  $M \times M$ .

Серьезной практической проблемой, связанной с реализацией ПУБФ на арифметике с фиксированной запятой, является потеря свойства перфективной реконструкции сигнала банком фильтров. Как показано ранее, это обусловлено чувствительностью структуры ПУБФ к квантованию коэффициентов. Решетчатые структуры четырех- и восьмиканальных ПУБФ на основе алгебры кватернионов, в отличие от известных решений, являются структурами без потерь даже при выполнении вычислений с конечной точностью. Например, для восьмиканального ПУБФ, характеристики которого показаны на рисунках 4.14 и 4.15, коэффициенты представлялись в формате с фиксированной запятой (таблица 4.1):

- а) шестизрядный формат слова, дополнительный код С2;
- б) восьмизрядный формат слова, кодирование CSD 8-2 – canonic signed digit (CSD  $m-n$ , где  $m$  – длина слова,  $n$  – число ненулевых битов).

Таблица 4.1 – Коэффициенты восьмиканального ПУБФ на основе алгебры кватернионов

Кватернион	Точное значение	Квантов.	Код CSD	Кватернион	Точное значение	Квантов.	Код CSD
$P_{0,1}$	-0,130700582784509	-0,132810	$-2^{-3} - 2^{-7}$	$P_{1,1}$	0,46521858651871	0,46875	$2^{-1} - 2^{-5}$
$P_{0,2}$	-0,077761503065683	-0,078125	$-2^{-4} - 2^{-6}$	$P_{1,2}$	0,44577292768800	0,43750	$2^{-1} - 2^{-4}$
$P_{0,3}$	-0,097626769294431	-0,093750	$-2^{-3} + 2^{-5}$	$P_{1,4}$	0,63512973485552	0,62500	$2^{-1} + 2^{-3}$
$P_{0,4}$	-0,983534198804438	-0,984380	$-1 + 2^{-6}$	$P_{1,4}$	-0,42599106047000	-0,43750	$-2^{-1} + 2^{-4}$
$Q_{0,1}$	0,914012880346929	0,937500	$1 - 2^{-4}$	$Q_{1,1}$	0,96157949993742	0,96875	$1 - 2^{-5}$
$Q_{0,2}$	0,170138307386757	0,156250	$2^{-3} + 2^{-5}$	$Q_{1,2}$	-0,13694209562823	-0,14063	$-2^{-3} - 2^{-6}$
$Q_{0,3}$	-0,313142675509949	-0,312500	$-2^{-2} - 2^{-4}$	$Q_{1,3}$	0,12972421500379	0,13281	$2^{-3} + 2^{-7}$
$Q_{0,4}$	0,193842914995494	0,187500	$2^{-2} - 2^{-4}$	$Q_{1,4}$	-0,19945765411912	-0,18750	$-2^{-2} + 2^{-4}$

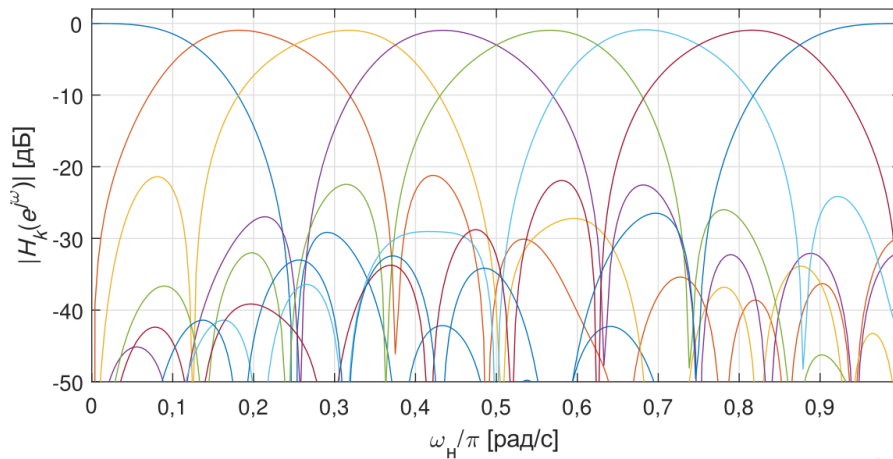
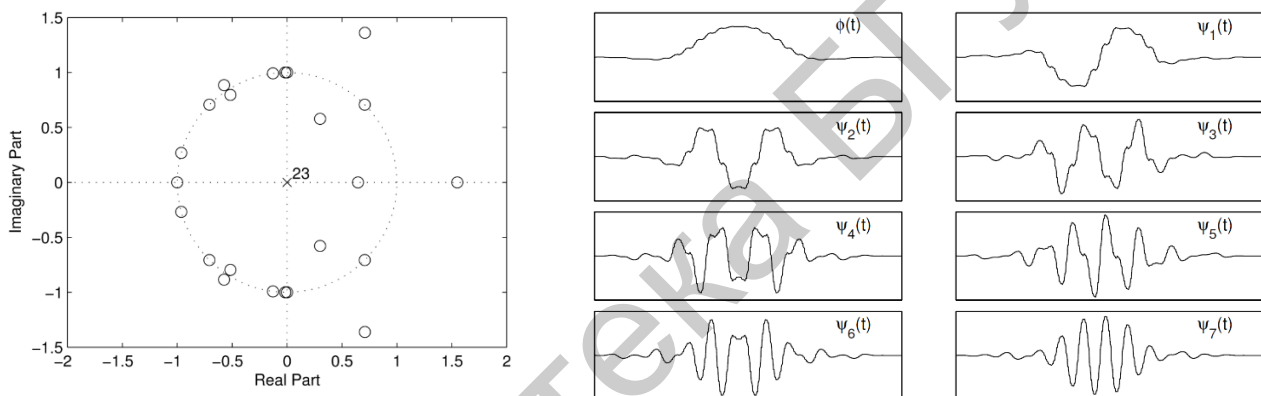


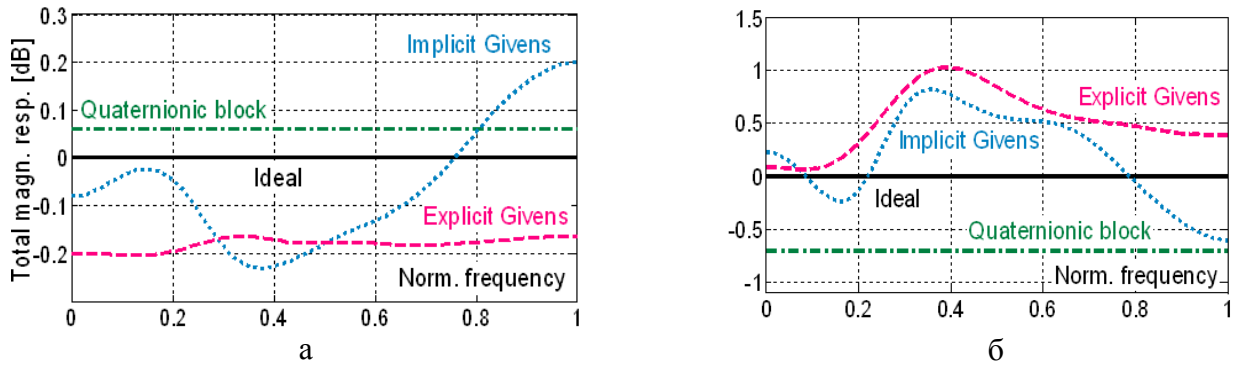
Рисунок 4.14 – Амплитудно-частотные характеристики канальных фильтров банка (ослабление в полосе непропускания – более 21 дБ, эффективность кодирования  $CG = 9,37$  дБ)



а – плоскость нулей низкочастотного фильтра  $H_0(z)$  ;  
 б – масштабная и вейвлет-функция

Рисунок 4.15 – Характеристики рассматриваемого банка фильтров

Эффект квантования коэффициентов ПУБФ анализировался для системы анализа-синтеза восьмиканального ПУБФ при трех схемах параметризации решетчатых структур банка фильтров на основе планарных оборотов Гивенса, лестничной параметризации матриц вращения Гивенса, а также матриц умножения кватернионов, т. е. для Q-ПУБФ:  $|T(e^{j\omega})| = \sum_{k=0}^7 |H_k(e^{j\omega})|^2$ . На рисунке 4.16 показаны АЧХ-системы анализа-синтеза восьмиканального ПУБФ для трех схем решетчатых структур банка фильтров: планарных оборотов Гивенса (explicit Givens), лестничной параметризации матриц вращения Гивенса (implicit Givens), а также Q-ПУБФ (quaternionic block).



а – шестиразрядный формат слова, дополнительный код C2;  
б – восьмиразрядный формат слова, кодирование CSD 8-2

Рисунок 4.16 – АЧХ-системы анализа-синтеза восьмиканального ПУБФ

Анализ результатов показывает, что решетчатые структуры ПУБФ на кватернионах являются системами без потерь, несмотря на квантование коэффициентов, частотная характеристика  $|T(e^{j\omega})|$  которых отличается от идеальной на некоторую постоянную величину. Корректирующий фактор определяется как значение  $|T(e^{j\omega})|$  для нулевой частоты:

$$c^2 = \sum_{k=0}^7 |H_k(e^{j\omega})|_{\omega=0}^2 = \sum_{k=0}^7 \left| \sum_{l=0}^{L-1} h_k(l) \right|^2.$$

Однако один из недостатков кватернионной факторизации заключается в большом количестве вычислений, что является ценой за сохранение ортогональности. Вычислительная сложность факторизаций (4.19) в два раза больше чем простое умножение матрицы  $4 \times 4$  на вектор, а (4.20) – в четыре раза больше. Однако, учитывая специфическую структуру матриц умножения кватернионов, можно выполнить данные операции только за восемь реальных умножений. Реализация умножителей кватернионов на основе распределенной арифметики или с использованием четырехмерных алгоритмов CORDIC хорошо согласуется со структурами FPGA.

Таким образом, умножитель кватернионов является альтернативным модулем построения ПУБФ и может рассцениваться как обобщение планарных вращений Гивенса. Решетчатые структуры ПУБФ на кватернионах являются системами без потерь, несмотря на квантование коэффициентов, и могут рассматриваться как расширение классических двухканальных решетчатых структур, разработанных Вайдианасаном 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. Q-ПУБФ могут рассматриваться как базис многополосных вейвлет-преобразований, эффективность которых в сжатии изображений высокая.

## 5 АЛГОРИТМЫ ВЫЧИСЛЕНИЯ ДКП

### 5.1 Применение ДКП в компрессии изображений

Наиболее часто при трансформационном кодировании применяется дискретное косинусное преобразование (ДКП) I, II, III, ..., VIII типов. ДКП-II чаще всего используется в данных системах. Математическая формула для ДКП-II длиной  $N$  определяется следующим образом:

$$X(k) = c(k) \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x(n) \cos\left(\frac{(2n+1)k\pi}{2N}\right), \quad (5.1)$$

$$\text{где } c(k) = \begin{cases} \frac{1}{\sqrt{N}}, & k = 0, \\ 1, & k > 0, \end{cases} \quad k = 0, 1, \dots, M-1;$$

$x(n)$  – оригинальный сигнал;

$X(k)$  – частотные коэффициенты исходного сигнала.

ДКП, как и преобразование Фурье, имеет обратное преобразование:

$$x(n) = c(k) \sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} X(k) \cos\left(\frac{(2n+1)k\pi}{2N}\right). \quad (5.2)$$

На практике в стандартах трансформационного кодирования, таких, как JPEG, MPEG, H.264 и др., используется двухмерное ДКП-II (2-D ДКП-II). Двухмерное дискретное косинусное преобразование для одного блока изображения размером  $M \times M$  вычисляется следующим образом:

$$\mathbf{B} = \mathbf{T} \cdot \mathbf{A} \cdot \mathbf{T}^T,$$

где  $\mathbf{A}$  – матрица преобразуемого блока изображения размером  $M \times M$  ( $N = 16$  или  $8$ );

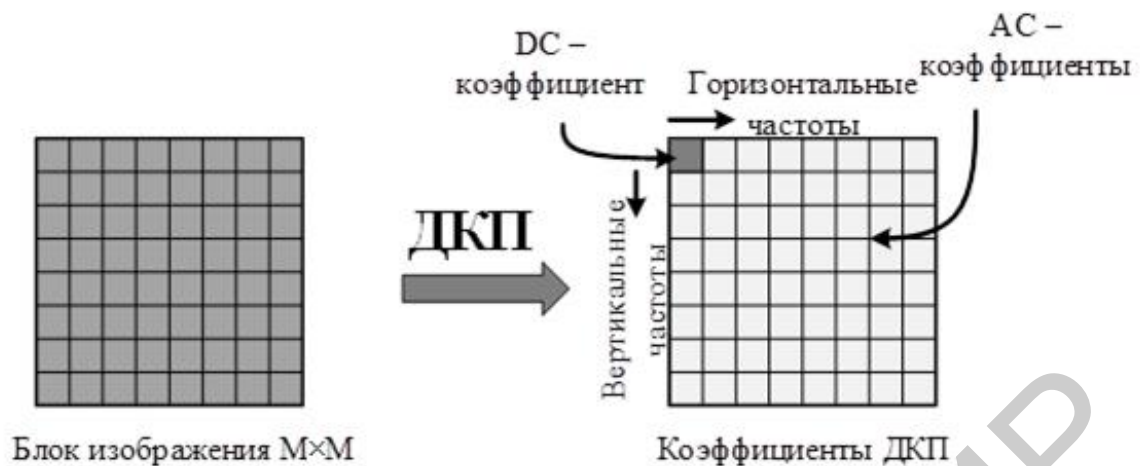
$\mathbf{T}$  – матрица коэффициентов ДКП-II;

$\mathbf{T}^T$  – транспонированная матрица коэффициентов ДКП-II.

Обратное двухмерное дискретное косинусное преобразование вычисляется как

$$\mathbf{A} = \mathbf{T}^T \cdot \mathbf{B} \cdot \mathbf{T}.$$

Широкая распространенность ДКП обусловлена его хорошей способностью производить уплотнение энергии изображения. Модуль ДКП выполняет отображение участка рисунка в частотную область, причем огромное значение имеет то, что наиболее важная часть информации об участке рисунка сосредоточена в левом верхнем углу массива (DC – Direct Current), что позволяет проводить гораздо более эффективное квантование и сжатие (или кодирование), чем в существовавших до того форматах (рисунок 5.1).



DC – нулевой спектральный коэффициент;  
 AC – множество высокочастотных коэффициентов

Рисунок 5.1 – Трансформационное кодирование

В качестве пояснения свойства «уплотнения энергии» рассмотрим пример вычисления двумерных дискретных косинусных преобразований в блоках с размерами  $8 \times 8$  элементов исходного изображения. При реконструкции изображения будем учитывать только 10 коэффициентов из каждого блока, остальные приравняем к нулю. Ниже приведена MATLAB-модель применения ДКП-II для сжатия изображения:

```
I = imread('cameraman.tif');
I = im2double(I);
D = dctmtx(8);
dct = @(block_struct) D * block_struct.data * D';
B = blockproc(I, [8 8], dct);
mask = [1 1 1 1 0 0 0 0
        1 1 1 0 0 0 0 0
          1 1 0 0 0 0 0 0
          1 0 0 0 0 0 0 0
          0 0 0 0 0 0 0 0
          0 0 0 0 0 0 0 0
          0 0 0 0 0 0 0 0
          0 0 0 0 0 0 0 0];
B2 = blockproc(B, [8 8], @(block_struct) mask .* block_struct.data);
invdct = @(block_struct) D' * block_struct.data * D;
I2 = blockproc(B2, [8 8], invdct);
imshow(I), figure, imshow(I2);
```

На рисунке 5.2 представлено два изображения – исходное и реконструированное. При реконструкции изображения использовалось только 15 % коэффициентов дискретных косинусных преобразований. Однако следует отметить, что качество реконструированного изображения является довольно приемлемым, но появляются артефакты, деградирующие декодированное изображение, в виде

«звона» или «ореола» на границах темного и светлого цветов, а также если при-  
 смотреться, то видны артефакты блочности, т. е. реконструированное изображе-  
 ние состоит из блоков или квадратиков  $8 \times 8$  пикселей.



а



б

а – оригинальное изображение; б – реконструированное изображение

Рисунок 5.2 – Пример работы MATLAB-функции dctmtx (8)

## 5.2 Основные типы и свойства ДКП

Дискретное косинусное преобразование (ДКП), как и подавляющее боль-  
 шинство других преобразований, используемых в обработке сигналов, является  
 линейным преобразованием. Это значит, что при таком преобразовании сохра-  
 няется величина скалярного произведения векторов. ДКП является вариантом  
 косинусного преобразования для вектора вещественных чисел. Также ДКП тесно  
 связано с дискретным преобразованием Фурье.

В общем виде линейное преобразование можно записать следующим обра-  
 зом:

$$y = \sum_{0 \leq l < n} t_{k,l} s_l, \quad (5.3)$$

где  $s = (s_0, \dots, s_{n-1})$  – входной сигнал;

$y = (y_0, \dots, y_{n-1})$  – выходной сигнал;

$t_{k,l}$  – коэффициенты преобразования.

Иногда более удобной записью является запись в векторно-матричной  
 форме:

$$y = \mathbf{T} s, \quad (5.4)$$

где  $\mathbf{T} = [t_{k,l}]_{0 \leq k, l < n}$ .



Дискретное преобразование Фурье (5.5) часто используется для обработки звука или подобных сигналов. Однако то, что ДПФ отображает вещественные значения в комплекснозначные не всегда удобно. Можно получить вещественное преобразование, если разложить комплексное ДПФ на его вещественную и мнимую части (5.6).

$$X_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x_n W_M^{-nk}, \quad (5.5)$$

$$(W_N)_{nk} = \cos\left(-\frac{2\pi nk}{N}\right) + i \sin\left(-\frac{2\pi nk}{N}\right). \quad (5.6)$$

Необходимо заметить, что ни синусная, ни косинусная части не являются полезными в качестве ядра преобразования, поскольку эти функции не образуют полный базис векторного пространства. Косинусные и синусные функции лишь стягивают подпространства четных и нечетных функций соответственно.

Эта проблема может быть решена с помощью ограничения косинусного преобразования положительным полупространством в пространственной области и в Фурье-области. В случае дискретных преобразований добавление тригонометрических функций с полуцелыми длинами волн может образовать базисные векторы с отсутствующей симметрией. Это эквивалентно удвоению базисной длины волны. Следовательно, ядро косинусного преобразования в  $N$ -мерном векторном пространстве есть

$$c_{nk} = \sqrt{\frac{2}{N}} \cos\left(\frac{\pi nk}{N}\right). \quad (5.7)$$

На рисунке 5.3 показаны базисные функции одномерных косинусоидных функций. Из-за удвоения периодов преобразование содержит четные и нечетные функции. Базисные функции с полуцелыми длинами волн замещают функции с исходно отсутствующей симметрией.

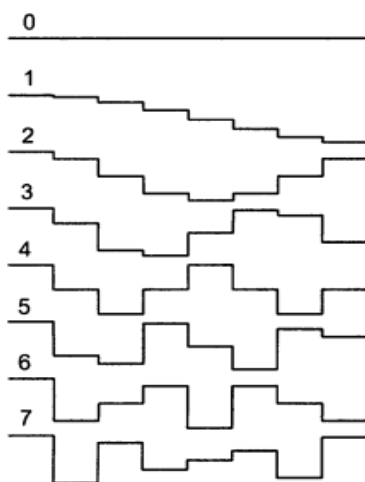


Рисунок 5.3 – Базисные функции одномерного унитарного пространства для  $N = 8$

Как и в случае с ДПФ, при выполнении преобразования мы должны дополнить входной вектор. В зависимости от выбранного способа периодического продолжения входного сигнала выделяют несколько основных типов ДКП:

$$\text{ДКП} - I_n = \left[ \cos kl \frac{\pi}{n-1} \right]_{0 \leq k, l < n}; \quad (5.8)$$

$$\text{ДКП} - II_n = \left[ \cos k \left( l + \frac{1}{2} \right) \frac{\pi}{n} \right]_{0 \leq k, l < n}; \quad (5.9)$$

$$\text{ДКП} - III_n = \left[ \cos \left( k + \frac{1}{2} \right) l \frac{\pi}{n} \right]_{0 \leq k, l < n}; \quad (5.10)$$

$$\text{ДКП} - IV_n = \left[ \cos \left( k + \frac{1}{2} \right) \left( l + \frac{1}{2} \right) \frac{\pi}{n} \right]_{0 \leq k, l < n}. \quad (5.11)$$

Если внимательно посмотреть на формулы (5.8)–(5.11), то можно заметить, что ДКП – III<sub>n</sub> (5.10) является обратным преобразованием для ДКП – II<sub>n</sub> (5.9).

Как видно, сигнал можно продолжать с использованием свойств четности или нечетности функции. Также различия имеются в выборе отсчета для проведения оси симметрии. Различные варианты продолжения сигнала показаны на рисунке 5.4.

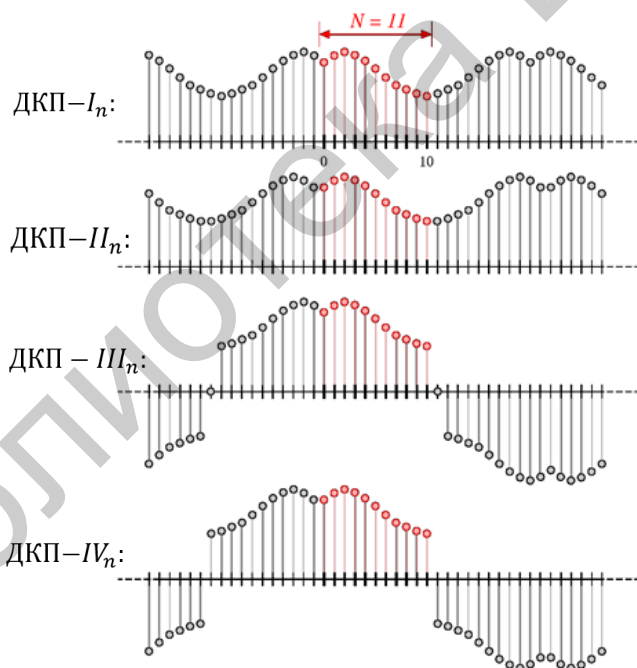


Рисунок 5.4 – Варианты дополнения входного вектора для наиболее распространенных типов ДКП

Необходимо также выделить основные свойства дискретного косинусного преобразования, которые играют важную роль при обработке сигналов, в частности изображений:

1) *Некоррелированность коэффициентов.* Коэффициенты независимы друг от друга, т. е. точность представления одного коэффициента не зависит от точности любого другого.

2) *«Уплотнение» энергии,* об этом говорилось выше. Преобразование сохраняет основную информацию в малом количестве коэффициентов. Данное свойство преобразования сыграло решающую роль в широком распространении ДКП.

3) *Сепарабельность, или делимость,* т. е. при обработке изображений вычисления производятся в два этапа путем последовательных операций ДКП по строкам и столбцам изображения.

4) *Симметричность.* Операции над строками и столбцами, так же как и формулы прямого и обратного преобразования ДКП, симметричны.

5) *Ортогональность.* Базисные функции ДКП ортогональны. Таким образом, обратная матрица преобразования ДКП равна транспонированной матрице ДКП.

### 5.3 Вычислительная сложность алгоритмов ДКП

Существует множество способов вычисления дискретного косинусного преобразования. Наиболее простой и очевидный способ – вычисление напрямую по одной из формул (5.8) – (5.11). Однако этот способ, естественно, является самым затратным, поэтому может использоваться лишь в качестве учебной иллюстрации. Гораздо более выигрышным способом является использование матрицы косинусного преобразования, которая может быть получена из формулы (5.7). Однако и такой способ все еще остается очень затратным.

Многие алгоритмы в цифровой обработке сигналов основаны на использовании линейных дискретных преобразований сигнала. Математически такое преобразование является умножением вектора на матрицу  $a \rightarrow M \cdot a$ , где  $a \in \mathbb{F}^n$  – это дискретный сигнал, а  $M \in \mathbb{F}^{n \times n}$  – преобразование над некоторым базовым полем  $\mathbb{F}$ . Решающим фактором применимости преобразования  $M$  является существование быстрого алгоритма, который позволяет снизить вычислительную сложность с  $O(n^2)$  операций, которые необходимы для умножения вектора на матрицу, до  $O(n \log(n))$  или менее операций. Задача нахождения этих алгоритмов для различных преобразований явилась важнейшей темой многих исследований и привела к появлению большого количества публикаций по обработке сигналов и математике.

Долгое время ученые не уделяли дискретному косинусному преобразованию особого внимания. Так, наиболее известный алгоритм быстрого преобразования Фурье был предложен Кули и Тьюки в 1965 году, а алгоритм быстрого косинусного преобразования был предложен на десятилетие позже.

Алгоритм быстрого ДКП был впервые представлен в 1974 году в работе Н. Ахмеда. В скором времени появились различные вариации данного алго-

ритма, авторы которых пытались уменьшить количество вычислений, необходимых для выполнения преобразования. Поскольку до 90-х годов операция умножения была очень затратной как при аппаратной реализации преобразования, так и при вычислении его на универсальном процессоре, то авторы старались уменьшить число умножений в алгоритме до теоретического минимума. Для иллюстрации можно привести таблицу, в которой показаны так называемые поколения алгоритмов ДКП.

Каждый из этих алгоритмов был получен путем сложных манипуляций с элементами матрицы преобразования. Данные алгоритмы обладают определенной структурой, поэтому их можно записать, как разложение матрицы преобразования в произведение слабозаполненных матриц, как будет показано в дальнейшем. Как видно из таблицы 5.1, наименьшее количество умножений используется в структуре Лофлера. Структура Лофлера является менее затратной по вычислительной сложности, поэтому будем ее использовать как основную схему при проектировании ДКП для встраиваемых мультимедийных систем.

Таблица 5.1 – Количество операций для восьмиточечного ДКП

Автор	Чен	Ван	Ли	Ветгерли	Сухиро	Лофлер
Число умножений	16	13	12	12	12	11
Число сложений	26	29	29	29	29	29

## 5.4 Алгоритмы вычисления ДКП

### 5.4.1 Алгоритм ДКП Ветгерли и Лайтенберга

Алгоритм был предложен в 1986 году. Авторы Ветгерли и Лайтенберг используют достаточно популярную схему «разделяй и властвуй» («Divide-and-Conquer»), немного ее видоизменив по сравнению с классическим вариантом. Введем следующие обозначения:

$$C_k = \cos \frac{k\pi}{16}, S_k = \sin \frac{k\pi}{16}, k = -7, \dots, +7, \quad (5.12)$$

тогда из-за симметричности функций синуса и косинуса мы получаем следующие равенства:

$$C_1 = S_7, C_2 = S_6, \dots, C_7 = S_1, \quad (5.13)$$

а также

$$C_1 = -C_{15}, C_2 = -C_{14}, \dots, C_7 = -C_9. \quad (5.14)$$

Обозначим различные варианты сумм и разностей элементов входного вектора  $f(n)_{n=0}^7$  следующим образом:

$$\begin{aligned}
s_{07} &= f(0) + f(7); \\
s_{12} &= f(1) + f(2); \\
s_{34} &= f(3) + f(4); \\
s_{56} &= f(5) + f(6); \\
s_{0734} &= s_{07} + s_{34}; \\
s_{1256} &= s_{12} + s_{56}; \\
s_{5634} &= d_{56} - s_{34}.
\end{aligned} \tag{5.15}$$

$$\begin{aligned}
d_{07} &= f(0) - f(7); \\
d_{12} &= f(1) - f(2); \\
d_{34} &= f(3) - f(4); \\
d_{56} &= f(5) - f(6); \\
d_{0734} &= s_{07} - s_{34}; \\
d_{1256} &= d_{12} - d_{56}; \\
\tilde{d}_{1256} &= d_{12} - s_{56}.
\end{aligned} \tag{5.16}$$

Тогда используя выражения (5.12)–(5.16), можно получить следующие формулы для дискретного косинусного преобразования:

$$\begin{aligned}
2\tilde{F}(0) &= C_4[(s_{07} + s_{12}) + (s_{34} + s_{56})]; \\
2\tilde{F}(4) &= C_4[(s_{07} + s_{34}) - (s_{12} + s_{56})]; \\
2\tilde{F}(2) &= C_{-6}[d_{12} - d_{56}] - S_{-6}[s_{07} - s_{34}]; \\
2\tilde{F}(6) &= S_{-6}[d_{12} - d_{56}] + C_{-6}[s_{07} - s_{34}]; \\
2\tilde{F}(1) &= C_1(d_{07} + C_4\tilde{d}_{1256}) + S_1(d_{34} + C_4s_{5634}); \\
2\tilde{F}(7) &= S_1(d_{07} + C_4\tilde{d}_{1256}) - C_1(d_{34} + C_4s_{5634}); \\
2\tilde{F}(3) &= C_3(d_{07} - C_4\tilde{d}_{1256}) - S_3(d_{34} - C_4s_{5634}); \\
2\tilde{F}(5) &= S_3(d_{07} - C_4\tilde{d}_{1256}) + C_3(d_{34} - C_4s_{5634}).
\end{aligned} \tag{5.17}$$

В алгоритме применяются операции поворота для входного вектора  $(x, y)$  на угол  $\Theta$ , при этом используются следующие выражения:

$$X = x \cos \Theta - y \sin \Theta, \tag{5.18}$$

$$Y = x \sin \Theta + y \cos \Theta. \tag{5.19}$$

При этом для угла поворота  $\Theta = \frac{k\pi}{16}$  выражения (5.18) и (5.19) можно переписать следующим образом:

$$X = Ck(x + y) - (Sk + Ck)y, \quad (5.20)$$

$$Y = C_k(x + y) + (S_k - C_k)x. \quad (5.21)$$

После некоторого упрощения и преобразования выражений (5.20) и (5.21) можно получить схему преобразования, показанную на рисунке 5.5. Блоки  $C_4$  обозначают умножение сигнала на константу  $C_4 = 1/\sqrt{2}$ . Блоки поворота поворачивают входной вектор на углы  $\frac{4\pi}{16}$ ,  $-\frac{6\pi}{16}$ ,  $\frac{\pi}{16}$  и  $\frac{3\pi}{16}$  соответственно. При этом один блок поворота содержит три умножения и три сложения, что дает 13 умножений и 29 сложений для схемы.

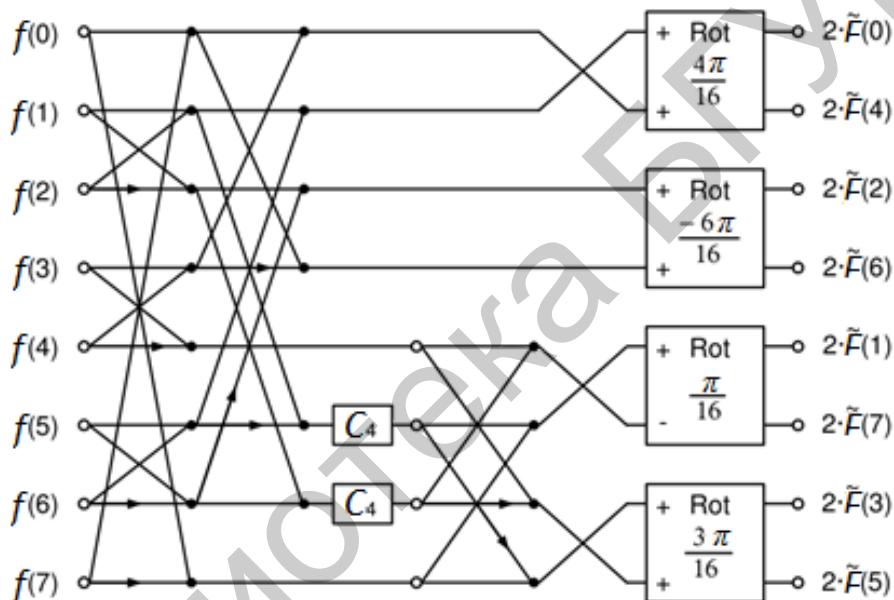


Рисунок 5.5 – Схема ДКП по Веттерли

#### 5.4.2 Алгоритм ДКП Лофлера

Лофлер предложил алгоритм восьмиточечного дискретного косинусного преобразования, который требует лишь 11 умножений и 29 сложений. На настоящий момент данный алгоритм преобразования является одним из наиболее распространенных благодаря малому количеству умножений. Схема преобразования довольно проста и представлена в матричном виде выражением (5.22).

$$\begin{pmatrix} F_0 \\ F_1 \\ F_2 \\ F_3 \\ F_4 \\ F_5 \\ F_6 \\ F_7 \end{pmatrix} = \frac{1}{2\sqrt{2}} \cdot \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & \sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{2} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sqrt{2}C_6 & \sqrt{2}S_6 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\sqrt{2}S_6 & \sqrt{2}C_6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & C_3 & 0 & 0 & S_3 \\ 0 & 0 & 0 & 0 & 0 & C_1 & S_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -S_1 & C_1 & 0 \\ 0 & 0 & 0 & 0 & -S_3 & 0 & 0 & C_3 \end{pmatrix} \times \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{pmatrix}. \tag{5.22}$$

Данная схема содержит поворачивающие блоки. После преобразования результат вычисления ДКП нуждается в масштабировании на  $2\sqrt{2}$ , что, однако, не сильно усложняет схему, так как после выполнения двумерного преобразования результат должен быть смасштабирован на 8, что можно выполнить простым сдвигом.

В алгоритме используются операции поворота для входного вектора  $(x, y)$  на угол  $\Theta$ , при этом используются выражения (5.18)–(5.21) и также можно получить схему преобразования, показанную на рисунке 5.5. Блоки поворота поворачивают входной вектор на углы  $\frac{6\pi}{16}, \frac{3\pi}{16}, \frac{\pi}{16}$  соответственно. При этом один блок поворота содержит три умножения и три сложения (рисунки 5.6, 5.7), что дает 11 умножений и 29 сложений для схемы.

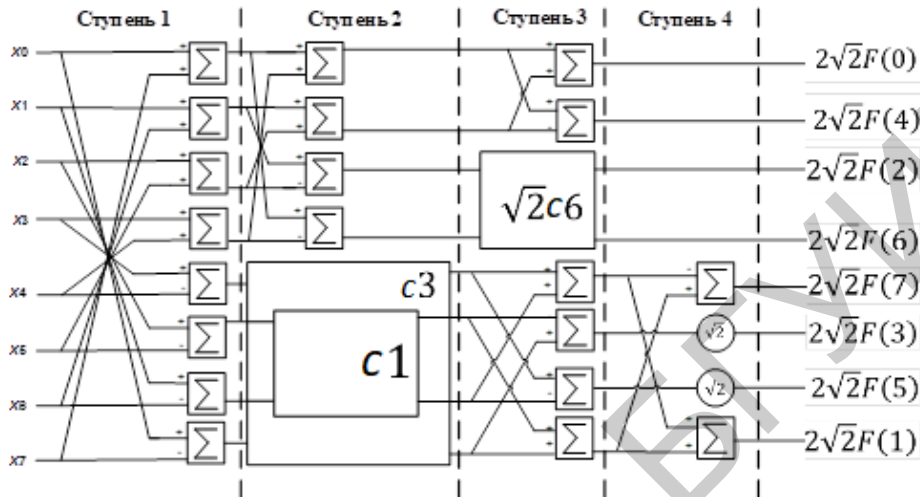


Рисунок 5.6 – Схема восьмиточечного ДКП II по Лофлеру с 11 умножениями

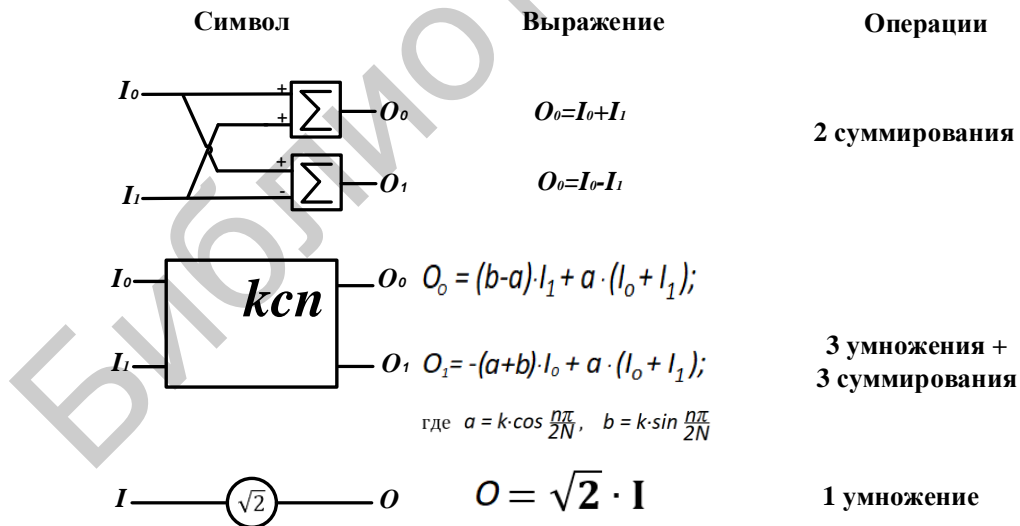


Рисунок 5.7 – Символы, используемые в структуре алгоритма Лофлера

В своей работе Лофлер показал, что алгоритмы ДКП можно разделить на несколько классов в зависимости от метода вычислений преобразования и по-



рядка выполнения поворотов. Так, классы, которые подразумевают параллельное выполнение умножений в рамках одной ступени всегда будут иметь как минимум 12 умножений. В то же время класс алгоритмов, которые используют последовательные умножения в рамках одной ступени, могут иметь на одно меньше. Также часть коэффициентов, выполняющихся на последней ступени, можно вынести в блок квантования, тем самым уменьшив количество операций умножения в блоке преобразования при трансформационном кодировании.

### 5.4.3 Алгоритм Линцера – Фейга

Линцер и Фейг описали алгоритм для двухмерного косинусного преобразования. Данный алгоритм также нашел широкое применение в сфере обработки аудио- и видеoinформации. Хотя авторами и описано двухмерное преобразование, сама схема представляет собой совокупность одномерных преобразований, применяемых последовательно к строкам и столбцам входного массива данных. Поэтому на рисунке 5.8 приведена схема одномерного преобразования.

В схеме присутствуют нетривиальные множители  $c_4$ ,  $t_1$ ,  $t_2$  и  $t_5$ , которые имеют следующие значения:

$$c_4 = \cos \frac{4\pi}{16}; t_1 = \operatorname{tg} \frac{\pi}{16}; t_2 = \operatorname{tg} \frac{2\pi}{16}; t_5 = \operatorname{tg} \frac{5\pi}{16}. \quad (5.23)$$

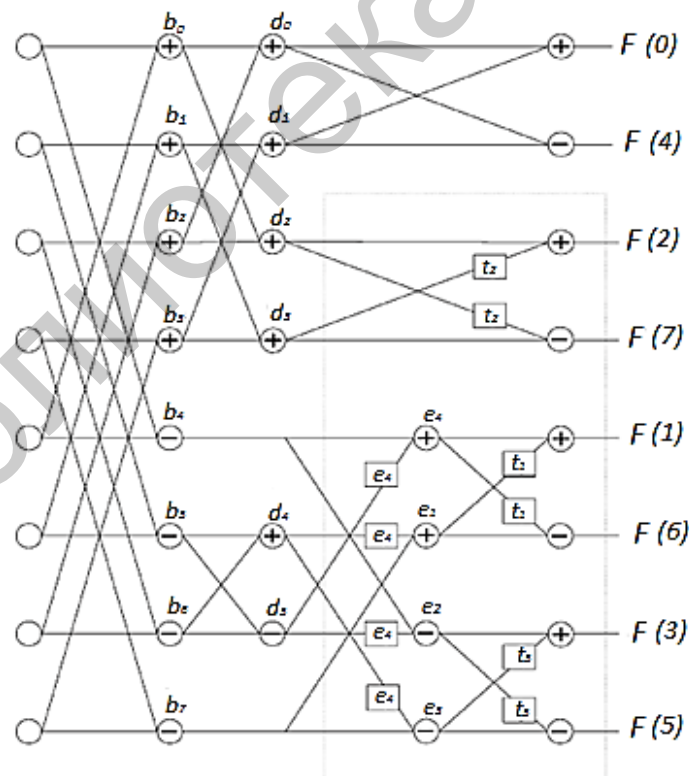


Рисунок 5.8 – Схема восьмиточечного ДКП по схеме Линцера – Фейга

## 5.5 Преобразователь ДКП – ОДКП на основе блочной лестничной структурной параметризации

Целочисленные ДКП (IntDCT), которые реализованы на основе лестничной схемной параметризации, были предложены для кодирования изображений по схеме L2L. Однако известные IntDCT генерируют в восстановленном исходном изображении такой артефакт, как «шахматная доска» при представлении коэффициентов преобразования в формате фиксированной запятой с малой разрядностью слова. В работе предложено преобразование для системы трансформационного кодирования на основе блочной лестничной структурной параметризации прямого и обратного ДКП (ДКП-ОДКП), где блоками лестничных ступеней являются матрицы прямого или обратного ДКП. Математическое описание и способы разложения прямого (рисунок 5.9, а, б) и обратного (рисунок 5.9, в, г) преобразования на основе параллельных двух диагональных преобразований  $\text{diag}\{\text{ДКП}, \text{ОДКП}\}$  следующие:

– прямого:

$$\begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{D} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C} & \mathbf{I} \end{bmatrix}; \quad (5.24)$$

$$\begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & -\mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{D} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{C} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{D} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}; \quad (5.25)$$

– обратного:

$$\begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{C} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{D} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{C} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{0} & -\mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix}; \quad (5.26)$$

$$\begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & -\mathbf{D} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{D} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{bmatrix}. \quad (5.27)$$

Прямое и обратное ДКП обозначены как матрицы  $\mathbf{C}$  и  $\mathbf{D}$  соответственно и функционально описываются как ДКП II типа и ДКП III типа:

$$[\mathbf{C}]_{m,n} = \sqrt{\frac{2}{M}} \cdot c_m \cos\left(\frac{m(n+\frac{1}{2})\pi}{M}\right), [\mathbf{D}]_{m,n} = \sqrt{\frac{2}{M}} \cdot c_n \cos\left(\frac{n(m+\frac{1}{2})\pi}{M}\right), \quad (5.28)$$

где  $\mathbf{I}$  – единичная матрица;

$$\mathbf{D} = \mathbf{C}^{-1} = \mathbf{C}^T, \quad 0 \leq m, n \leq M-1;$$

$[\mathbf{C}]_{m,n}$  – матрица прямого ДКП II типа;

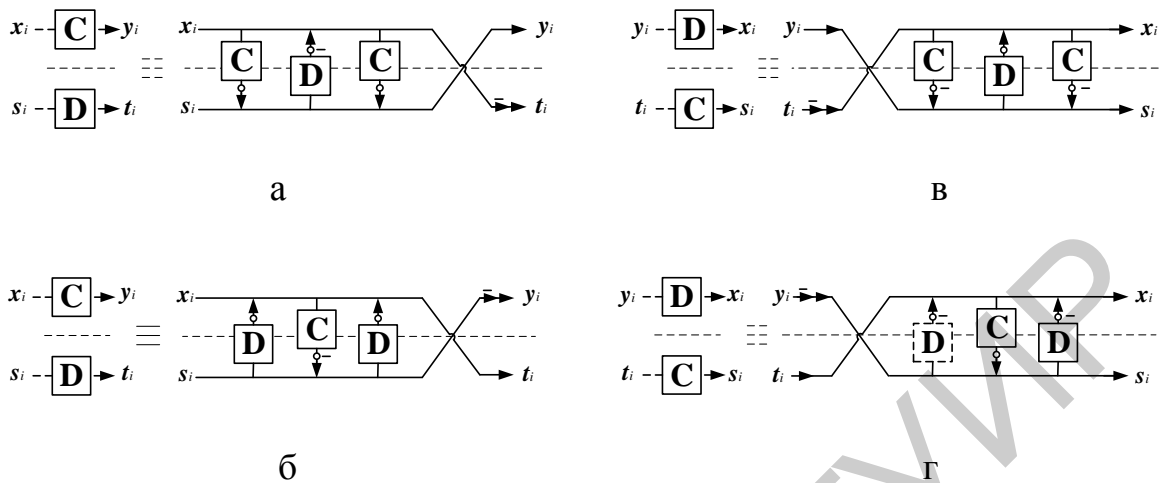
$[\mathbf{D}]_{m,n}$  – матрица ДКП III типа, являющаяся обратной для ДКП II типа;

$m$  – количество столбцов,

$n$  – количество строк;

$$M = 2^n \quad (n \in \mathbb{N}),$$

$$c_m = \begin{cases} \frac{1}{\sqrt{2}} (m=0) \\ 1 (m \neq 0) \end{cases}, \quad c_n = \begin{cases} \frac{1}{\sqrt{2}} (n=0) \\ 1 (n \neq 1) \end{cases}.$$



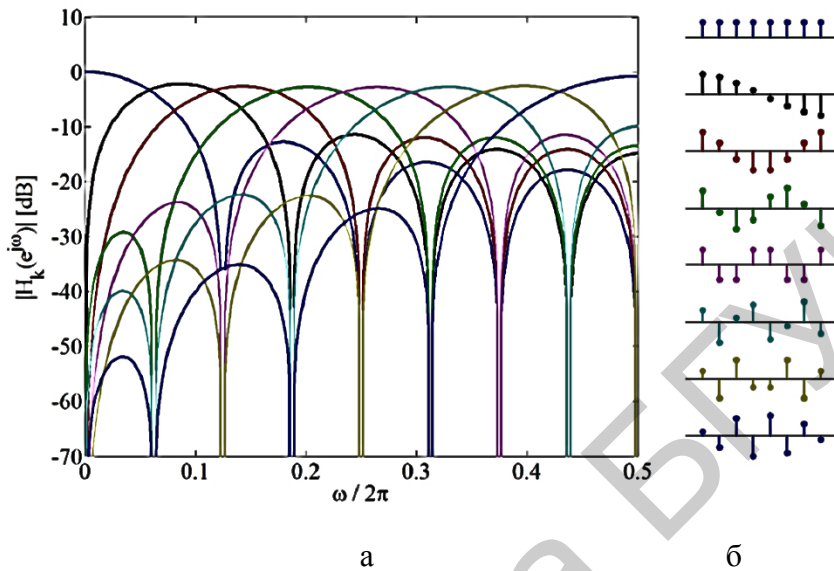
- а – вариант I прямого преобразования на основе ДКП – ОДКП;
- б – вариант II прямого преобразования на основе ДКП – ОДКП;
- в – вариант I обратного преобразования на основе ДКП – ОДКП;
- г – вариант II обратного преобразования на основе ДКП – ОДКП

Рисунок 5.9 – Варианты преобразований на основе блочной лестничной структурной параметризации параллельных диагональных преобразований ДКП – ОДКП (пустые кружки на схеме обозначают округление)

Амплитудно-частотные и импульсные характеристики восьмиточечного прямого и обратного ДКП представлены на рисунках 5.10 и 5.11 соответственно. Анализ рисунков показывает, что преобразование ДКП является равнополосным максимально децимированным банком фильтров с линейной ФЧХ (импульсные характеристики обладают свойством симметрии), постоянная составляющая локализована первым фильтром банка. Однако все фильтры ОДКП имеют постоянную составляющую (рисунок 5.11, а), т. е. присутствует так называемый эффект DC-leakages.

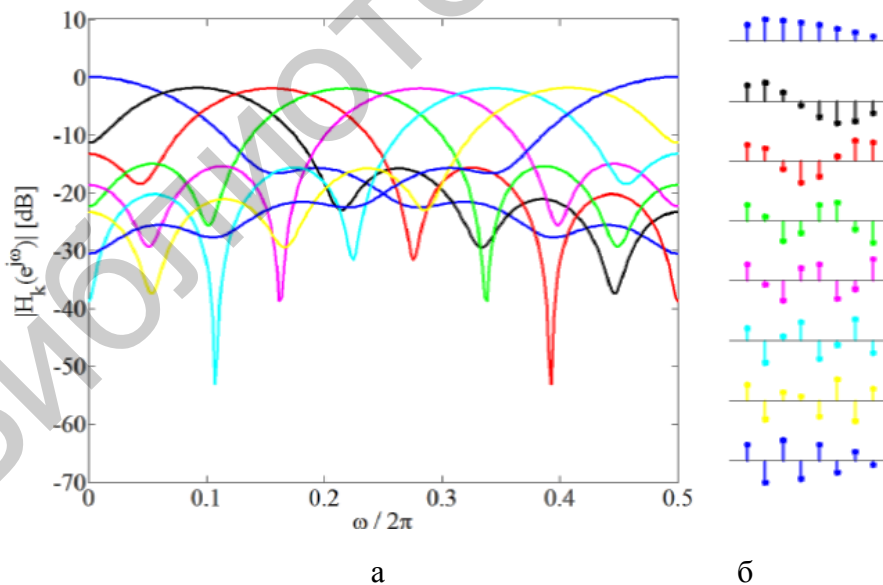
В программной среде MATLAB с помощью инструментального модуля Fixed Point для моделирования вычислений на арифметике с фиксированной запятой был осуществлен расчет амплитудно-частотных характеристик (АЧХ) системы анализа-синтеза прямого – обратного преобразования ДКП – ОДКП с блочной лестничной структурной параметризацией и преобразования ДКП – ОДКП стандарта JPEG для разрядности данных 12 бит, из которых 1 бит знаковый, 1 бит – целая часть слова, оставшиеся 10 бит представляют дробную часть слова. Анализ АЧХ данных преобразований (рисунок 5.12) показывает, что АЧХ системы анализа-синтеза преобразования ДКП – ОДКП с блочной лестничной структурной параметризацией (рисунок 5.12, а) представляет собой прямолинейную АЧХ, т. е. система анализа-синтеза данного преобразования вносит только задержку. Между тем АЧХ (рисунок 5.12, б) системы анализа-синтеза прямого и

обратного ДКП стандарта JPEG соответствует нелинейной функции с максимальной неравномерностью, равной  $-0,03$  дБ. Таким образом, в отличие от алгоритма ДКП, используемого в стандарте сжатия изображений JPEG, преобразование на основе блочной лестничной структурной параметризации прямого и обратного ДКП (ДКП – ОДКП) обладает свойством перфективной реконструкции исходного изображения.



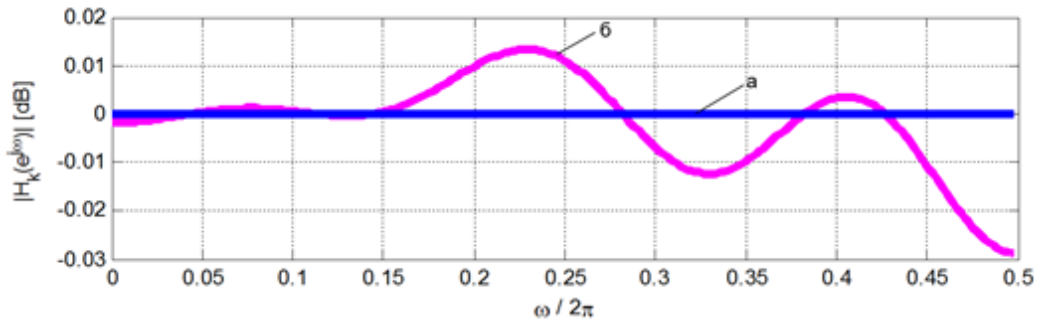
а – амплитудно-частотные; б – импульсные

Рисунок 5.10 – Характеристики прямого ДКП



а – амплитудно-частотные; б – импульсные

Рисунок 5.11 – Характеристики обратного ДКП



а – преобразование ДКП – ОДКП с блочной лестничной структурной параметризацией; б – ДКП – ОДКП стандарта JPEG

Рисунок 5.12 – АЧХ системы анализа-синтеза

## 5.6 Преобразование ДКП – ОДКП с совершенной реконструкцией

Проявление эффекта DC leakages (утечки энергии нулевой частоты в высокочастотные составляющие) в системе преобразования ДКП – ОДКП с блочной лестничной структурной параметризацией, обусловленного наличием постоянной составляющей во всех фильтрах ОДКП (см. рисунок 5.11, а), приводит к потере свойства регулярности первого рода<sup>1</sup> банком фильтров ДКП – ОДКП, и в реконструированном исходном изображении проявляется эффект «шахматной доски». Для устранения данного недостатка в работе предлагается преобразователь системы трансформационного кодирования по схеме L2L на основе ДКП – ОДКП с использованием дополнительного информационного SIB-блока (side information block)  $\mathbf{S}_i$  размером  $M \times M$ . Изначально  $s_0$  является нулевой матрицей, а  $s_i$  – итеративное преобразование от  $s_0$  до  $s_i$  как

$$s_i = \mathbf{D}s_{i-1} \text{ для } i = 1, 2, \dots, n-1. \quad (5.29)$$

Формулы для прямого и обратного преобразований имеют следующий вид:

$$\begin{bmatrix} y_0 \\ \vdots \\ y_{n-1} \\ s_n \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \cdots & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \mathbf{C} & \vdots \\ 0 & \cdots & \cdots & \mathbf{D}^n \end{bmatrix} \begin{bmatrix} x_0 \\ \vdots \\ x_{n-1} \\ s_0 \end{bmatrix}, \quad \begin{bmatrix} x_0 \\ \vdots \\ x_{n-1} \\ s_0 \end{bmatrix} = \begin{bmatrix} \mathbf{D} & \cdots & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \mathbf{D} & \vdots \\ 0 & \cdots & \cdots & \mathbf{C}^n \end{bmatrix} \begin{bmatrix} y_0 \\ \vdots \\ y_{n-1} \\ s_n \end{bmatrix}, \quad (5.30)$$

где  $\mathbf{C}$  и  $\mathbf{D}$  – матрицы прямого и обратного ДКП;

$s_n \neq 0$  из-за ошибки округления в каждом лифтинг-шаге.

Таким образом, можно сделать выводы:

1) В режиме без потерь изображение реконструируется из всего битового потока и информационной части  $s_n$ . Каждый блок  $s_i$  – это последовательно инвертируемые преобразования без любых потерь.

<sup>1</sup> Система отвечает условию регулярности первого рода (первого порядка), когда значения амплитудно-частотных характеристик всех канальных фильтров, кроме первого, равны нулю на нулевой частоте.



фиксированной запятой для исходного изображения Lena 512×512 пикселей (рисунок 5.15, а) показало, что без применения информационного SIB-блока  $s_n$  наблюдается эффект шахматной доски: PSNR = 84,6834 дБ, MSE = 2,2118×10<sup>-4</sup> (рисунок 5.15, б), а с обработкой по схеме с SIB-блоком  $s_n$  PSNR = ∞ дБ, MSE = 0, т. е. исходное изображение восстанавливается без потерь (рисунок 5.15, в).



а – исходное изображение; б – реконструированное изображение без применения информационного SIB-блока; в – реконструированное изображение с применением информационного SIB-блока

Рисунок 5.15 – Реконструкция изображения Lena 512×512 пикселей системой анализа-синтеза

Двухмерное ДКП для блока  $x$  размером  $M \times M$  описывается выражением

$$y = (C(Cx)^T)^T = Cx C^T, \quad (5.31)$$

где  $C$  – матрица прямого преобразования ДКП;

$C^T$  – транспонированная матрица, или обратное преобразование ДКП.

Математическое описание двухмерной лестничной структурной параметризации для прямого (рисунок 5.16, а, б) и обратного разложений (рисунок 5.16, в, г) будет следующее (формулы (5.24) – (5.27) соответственно):

– прямого двухмерного:

$$\begin{bmatrix} C_{2D} & 0 \\ 0 & D_{2D} \end{bmatrix} = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} \begin{bmatrix} I & 0 \\ C_{2D} & I \end{bmatrix} \begin{bmatrix} I & -D_{2D} \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ C_{2D} & I \end{bmatrix}; \quad (5.32)$$

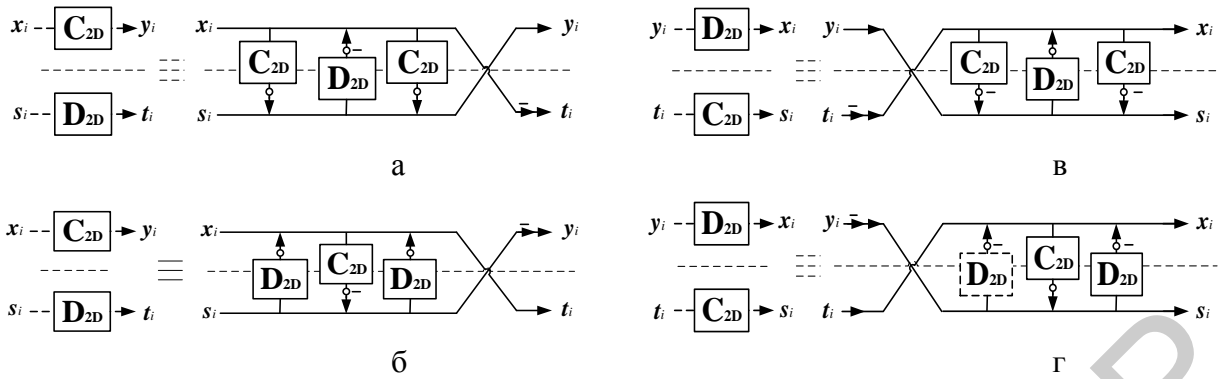
$$\begin{bmatrix} C_{2D} & 0 \\ 0 & D_{2D} \end{bmatrix} = \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix} \begin{bmatrix} I & D_{2D} \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ -C_{2D} & I \end{bmatrix} \begin{bmatrix} I & D_{2D} \\ 0 & I \end{bmatrix}; \quad (5.33)$$

– обратного двухмерного:

$$\begin{bmatrix} C_{2D} & 0 \\ 0 & D_{2D} \end{bmatrix}^{-1} = \begin{bmatrix} D_{2D} & 0 \\ 0 & C_{2D} \end{bmatrix} = \begin{bmatrix} I & 0 \\ -C_{2D} & I \end{bmatrix} \begin{bmatrix} I & D_{2D} \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ -C_{2D} & I \end{bmatrix} \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix}; \quad (5.34)$$

$$\begin{bmatrix} C_{2D} & 0 \\ 0 & D_{2D} \end{bmatrix}^{-1} = \begin{bmatrix} D_{2D} & 0 \\ 0 & C_{2D} \end{bmatrix} = \begin{bmatrix} I & -D_{2D} \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ C_{2D} & I \end{bmatrix} \begin{bmatrix} I & -D_{2D} \\ 0 & I \end{bmatrix} \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}; \quad (5.35)$$

где  $\mathbf{C}_{2D} \mathbf{x}_i \triangleq (\mathbf{C}(\mathbf{C}\mathbf{x}_i)^T)^T = \mathbf{C}\mathbf{x}_i\mathbf{C}^T$  и  $\mathbf{D}_{2D} \mathbf{x}_i \triangleq (\mathbf{D}(\mathbf{D}\mathbf{x}_i)^T)^T = \mathbf{D}\mathbf{x}_i\mathbf{D}^T$ .



а – вариант I прямого преобразования; б – вариант II прямого преобразования;  
в – вариант I обратного преобразования; г – вариант II обратного преобразования

Рисунок 5.16 – Схема двумерного преобразования для системы трансформационного кодирования на основе блочной лестничной структурной параметризации параллельных диагональных преобразований ДКП – ОДКП (пустые кружки на схеме обозначают округление)

Формулы для прямого и обратного двумерного преобразований с информационным СИВ-блоком будут иметь следующий вид согласно выражению (5.30):

$$\begin{bmatrix} y_0 \\ \vdots \\ y_{n-1} \\ s_n \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{2D} & \cdots & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \mathbf{C}_{2D} & \vdots \\ 0 & \cdots & \cdots & \mathbf{D}_{2D}^n \end{bmatrix} \begin{bmatrix} x_0 \\ \vdots \\ x_{n-1} \\ s_0 \end{bmatrix}; \begin{bmatrix} x_0 \\ \vdots \\ x_{n-1} \\ s_0 \end{bmatrix} = \begin{bmatrix} \mathbf{D}_{2D} & \cdots & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \mathbf{D}_{2D} & \vdots \\ 0 & \cdots & \cdots & \mathbf{C}_{2D}^n \end{bmatrix} \begin{bmatrix} y_0 \\ \vdots \\ y_{n-1} \\ s_n \end{bmatrix}. \quad (5.36)$$

Пример факторизации обратного двумерного преобразования ДКП – ОДКП на основе блочной лестничной структурной параметризации с использованием информационного СИВ-блока согласно формулам (5.35) – (5.36) показан на рисунке 5.17.

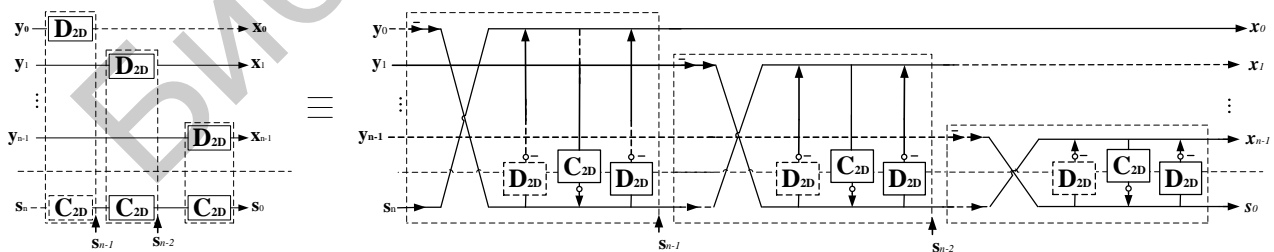


Рисунок 5.17 – Обратное двумерное преобразование ДКП – ОДКП



## 6 ПРИМЕНЕНИЕ МЕТОДОВ ТРАНСФОРМАЦИОННОГО КОДИРОВАНИЯ ДЛЯ ОБРАБОТКИ ИЗОБРАЖЕНИЙ

### 6.1 Применение Q-ПУБФ в компрессии изображений

С целью исследования возможностей Q-ПУБФ в системах сжатия цифровых изображений в режиме lossy (кодирования с потерями) и сравнения полученных результатов с системами на основе ПУБФ, БОБФ, многоуровневыми ДВП и другими ортогональными преобразованиями синтезирован восьмиканальный Q-ПУБФ с линейной ФЧХ и регулярностью первого порядка, коэффициенты которого приведены в таблице 6.1. Число ступеней декомпозиции передаточных матриц анализа  $\mathbf{E}(z)$  и синтеза  $\mathbf{D}(z)$  равно  $N = 3$ , длина импульсных характеристик канальных фильтров составляет 24 отсчета:  $8 \times 24$  Q-ПУБФ. Коэффициент эффективности кодирования данного банка фильтров равен  $CG = 9,3747$  дБ. Как видно из анализа амплитудно-частотной характеристики банка фильтров (см. рисунок 6.1), минимальное ослабление в полосе затухания составляет величину  $-21$  дБ, эффект DC leakage отсутствует.

Таблица 6.1 – Коэффициенты восьмиканального Q-ПУБФ с линейной ФЧХ,  $N = 3$  ( $8 \times 24$  Q-ПУБФ)

Коэффициенты	Re( $\cdot$ )	Im <sub>i</sub> ( $\cdot$ )	Im <sub>j</sub> ( $\cdot$ )	Im <sub>k</sub> ( $\cdot$ )
$s_0$	-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

Моделирование 2D трансформационного преобразования на основе восьмиполосного Q-ПУБФ ( $8 \times 24$  Q-ПУБФ) с блочной лестничной схемой параметризацией осуществлялось в среде MATLAB Fixed Point Toolbox. В качестве альтернативных использовались два восьмиканальных ПУБФ с длиной импульсной характеристики 16 отсчетов ( $8 \times 16$  FUFB) и 24 отсчета ( $8 \times 24$  FUFB) с показателем эффективности кодирования  $CG = 9,35$  дБ и  $9,49$  дБ соответственно, а также два восьмиканальных БОБФ с такой же длиной импульсных характеристик:  $8 \times 16$  BOFB ( $CG = 9,62$  дБ) и  $8 \times 24$  BOFB ( $CG = 9,68$  дБ). Данные банки фильтров являются банками с решетчатой параметризацией на основе планарных вращений Гивенса и регулярностью первого порядка. Для сравнения выбраны также восьмиканальный ПУБФ  $8 \times 16$  LOT, а также для ПУБФ и БОБФ с линейной ФЧХ соответственно  $8 \times 24$  GenLOT и  $8 \times 16$  GLTB,  $8 \times 24$  GLTB.

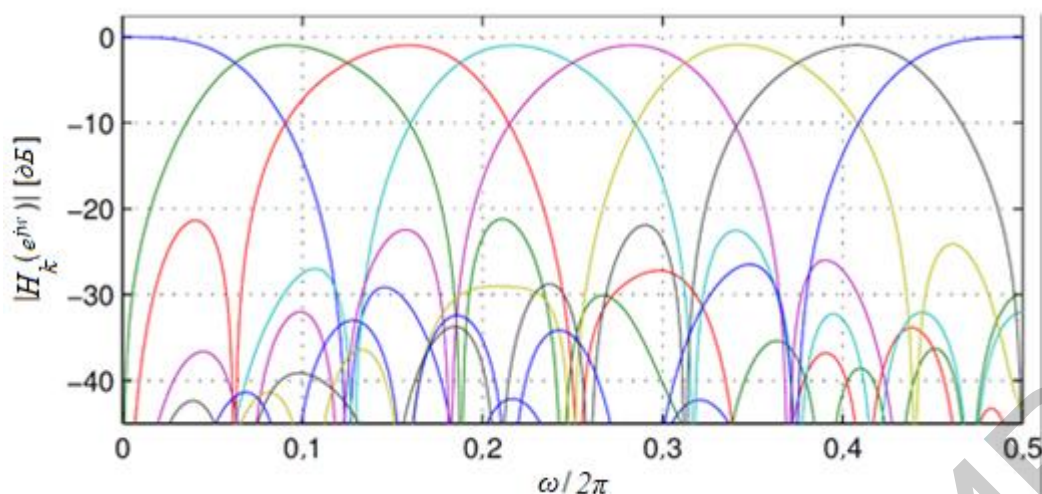


Рисунок 6.1 – АЧХ  $8 \times 24Q$ -ПУБФ с линейной ФЧХ,  $N = 3$

В оценке эффективности кодирования изображений по схеме lossy для обработки исходного изображения в 64 субполосах применялась шестиуровневая декомпозиция дискретного вейвлет-преобразования (Daubechies wavelet 12 с аппроксимацией линейной фазы). Метод нулевого дерева EZW-IP использовался для преобразования коэффициентов фильтрации и последующего квантования по методу оптимального распределения битов в параунитарном субполосном кодере изображений на основе  $Q$ -ПУБФ.

В таблице 6.2 и на рисунке 6.2 показаны сравнительные результаты кодирования двух восьмибитовых тестовых изображений  $512 \times 512$  «Lena» и «Barbara» по объективному показателю PSNR. Для сравнения также приведены показатели эффективности кодера JPEG2000 (9/7 wavelet) JASPER.

Таблица 6.2 – Сравнительные результаты кодирования восьмибитового тестового изображений  $512 \times 512$  «Lena» и «Barbara» по показателю PSNR

Тип банка фильтров	«Lena»			«Barbara»		
	0,25 bpp	0,5 bpp	1,0 bpp	0,25bpp	0,5bpp	1,0bpp
Daubechies wavelet 12	33,65	36,77	40,01	28,16	32,09	37,30
9/7 wavelet	33,71	36,83	40,03	28,10	32,01	37,17
$8 \times 16$ PUBF	33,17	36,57	39,73	29,20	33,31	38,30
$8 \times 16$ LOT	32,91	36,13	39,28	29,05	33,04	37,84
$8 \times 24$ PUBF	33,36	36,64	39,94	29,43	33,53	38,34
$8 \times 24$ GenLOT	33,25	36,54	39,82	29,31	33,55	38,31
$8 \times 16$ BOFB	33,43	36,67	39,73	29,31	33,33	38,26
$8 \times 16$ GLBT	33,35	36,62	39,70	29,23	33,28	38,19
$8 \times 24$ BOFB	33,53	36,82	39,84	29,66	33,63	38,38
$8 \times 24$ GLBT	33,32	36,61	39,68	29,29	33,29	38,18
$8 \times 24Q$ -ПУБФ	34,65	37,15	39,41	30,58	34,51	38,19
JPEG2000 (9/7 wavelet)	33,25	36,29	39,25	27,73	31,41	36,56
DCT-Loefflera(CSD8-4)	32,13	35,48	38,84	22,35	24,78	27,19

Анализ таблиц 6.2 и 6.3 показывает, что использование 2D-трансформационного преобразования на основе восьмиполосного  $Q$ -ПУБФ ( $8 \times 24Q$ -ПУБФ) в кодерах цифровых изображений по эффективности сопоставимо с другими из-

вестными преобразованиями, а благодаря большому количеству степеней свободы при синтезе данных банков фильтров преобразование  $8 \times 24Q$ -ПУБФ имеет выше эффективность по показателю PSNR для изображений с относительно сильными высокочастотными компонентами (см. таблицу 6.2). С другой стороны, субъективное сравнение восстановленных изображений «Lena» и «Barbara» для  $bpp$ : 0,25; 0,5; 1,0 (см. рисунок 6.3) показывает отсутствие характерных артефактов, например, эффекта Гиббса, волнообразных ложных контуров. При этом мелкие детали хорошо сохранились.

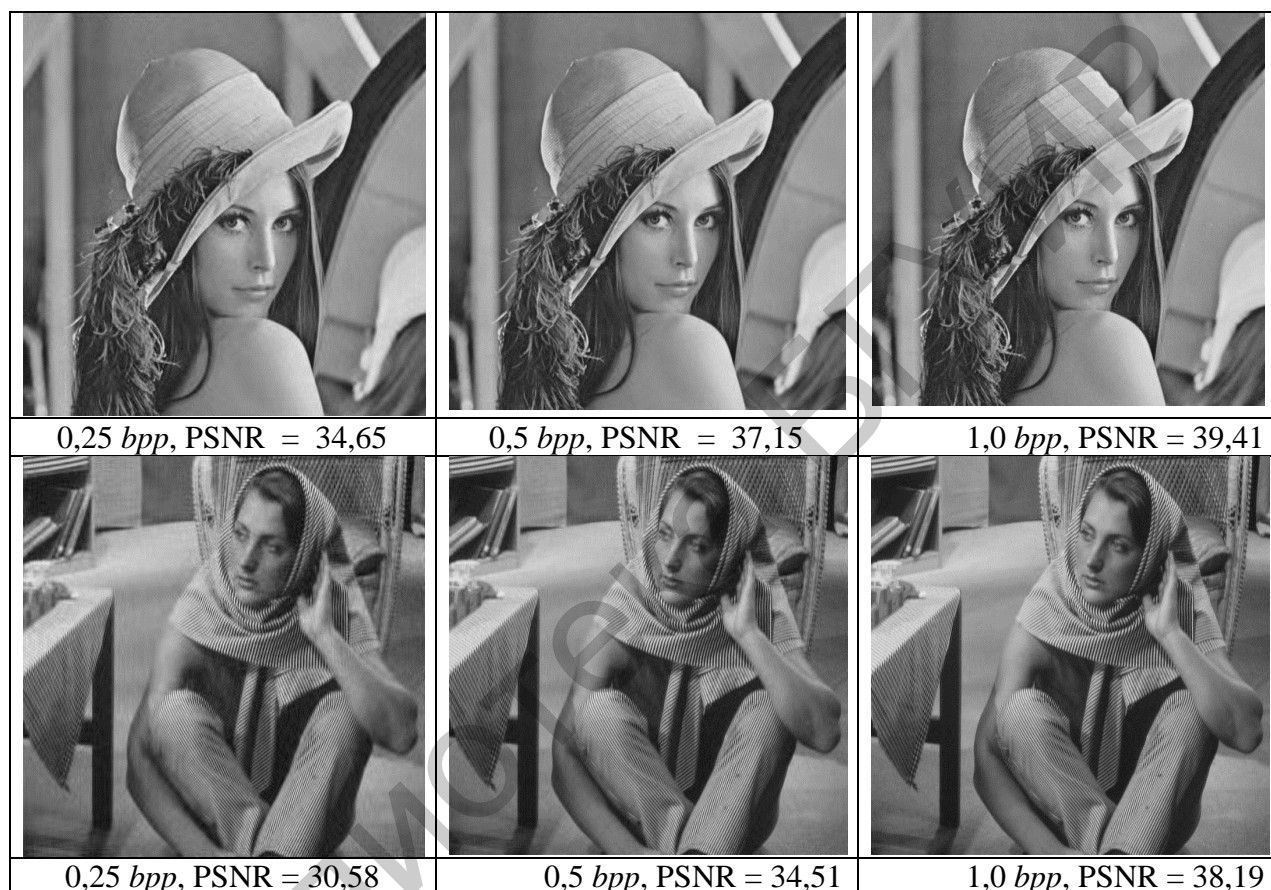


Рисунок 6.2 – Восстановленные изображения «Lena» и «Barbara» для разных значений среднего числа битов на пиксель  $bpp$

## 6.2 Компрессия изображений на основе ДКП – ОДКП

Для режима кодирования данных с потерями на основе схемы ДКП – ОДКП используются три цикла кодера  $L2L$  и система компрессии будет работать следующим образом:

- исходное изображение  $M \times M$  разбивается на блоки  $N \times 2N$ ;
- далее изображение разбивается на два изображения размером  $M/2 \times M$ , содержащих блоки  $x_i$  и  $s_i$  размером  $N \times N$ ;
- далее над блоками  $x_i$  и  $s_i$  производится прямая двухмерная блочная лестничная структурная параметризация;
- преобразованное изображение собирается по частотным каналам;

- собранное по частотам преобразованное изображение квантуется;
- далее происходит обратная процедура разборки по частотным каналам;
- применяется обратная двухмерная блочная лестничная структурная параметризация для квантованных коэффициентов;
- сборка реконструированного изображения  $M \times M$  из декодированных двух независимых изображений  $M/2 \times M$ , содержащих блоки  $y_i$  и  $t_i$ .

Как видно из таблицы 6.3 и рисунка 6.3, предлагаемая система компрессии на основе ДКП – ОДКП преобразователя, квантователя с оптимальным распределением битов по частотным субполосам и кодирования RLE+Huffman дает лучшие показатели PSNR на 0,5 bpp (сжатие 1:16) и 1 bpp (сжатие 1:8).

Таблица 6.3 – Сравнение аналогичных систем сжатия по показателю PSNR [дБ]

Image 512×512	Bit rate [bpp]	Предлагаемое DCT-IDCT в режиме lossy	Suzuki [75]	Komatsu [54]	Fuku- ma [44]	Tran [56]	Chokchaitam [35]
«Lena»	0,25	28,79	<b>31,86</b>	31,83	31,40	29,03	31,80
	0,50	<b>35,78</b>	35,56	34,38	34,41	32,25	35,40
	1,00	<b>43,96</b>	39,12	36,68	38,70	35,52	38,82
«Barbara»	0,25	25,12	<b>26,95</b>	26,93	26,94	23,70	26,69
	0,50	<b>32,42</b>	30,68	30,65	30,67	27,14	30,32
	1,00	<b>39,22</b>	36,03	35,88	35,97	31,18	35,73
«Baboon»	0,25	21,35	<b>22,57</b>	22,56	<b>22,57</b>	20,71	22,25
	0,50	<b>25,01</b>	24,81	24,80	24,80	21,85	24,59
	1,00	<b>37,93</b>	28,27	28,24	28,25	24,54	28,06
«Goldhill»	0,25	27,04	<b>29,37</b>	29,35	29,36	27,35	29,06
	0,50	<b>32,04</b>	31,97	31,92	31,94	29,27	31,74
	1,00	<b>46,88</b>	35,36	35,19	35,27	31,31	35,13




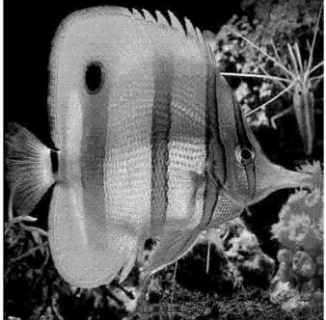
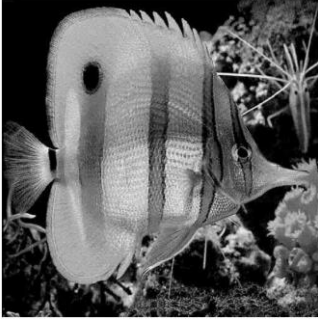
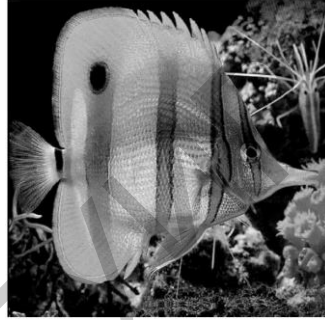



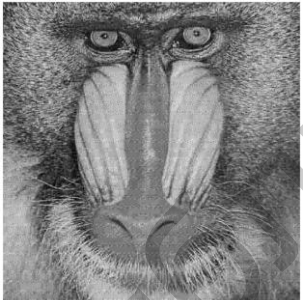
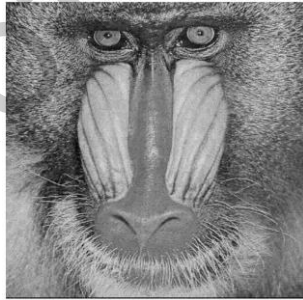
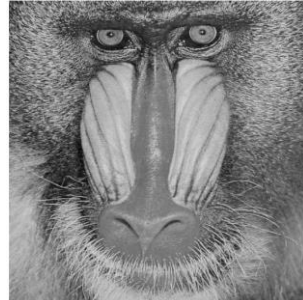



		
0,25 bpp, PSNR = 25,12 дБ	0,5 bpp, PSNR = 32,42 дБ	1 bpp, PSNR = 39,22 дБ
		
0,25 bpp, PSNR = 24,42 дБ	0,5 bpp, PSNR = 28,31 дБ	1 bpp, PSNR = 39,23 дБ
		
0,25 bpp, PSNR = 28,79 дБ	0,5 bpp, PSNR = 35,78 дБ	1 bpp, PSNR = 43,96 дБ
		
0,25 bpp, PSNR = 21,35 дБ	0,5 bpp, PSNR = 25,01 дБ	1 bpp, PSNR = 37,93 дБ
		
0,25 bpp, PSNR = 27,04 дБ	0,5 bpp, PSNR = 32,04 дБ	1 bpp, PSNR = 46,88 дБ

Рисунок 6.3 – Восстановленное изображение «Barbara», «Fish», «Lena», «Baboon», «Goldhill» при кодировании двумерным ДКП – ОДКП – квантователь – RLE+Huffman

## ЛИТЕРАТУРА

### **Цифровая обработка сигналов**

- 1 Голд, Б. Цифровая обработка сигналов / Б. Голд, Ч. Рэйдер. – М. Сов. радио, 1973. – 368 с.
- 2 Рабинер, Л. Теория и применение цифровой обработки сигналов / Л. Рабинер, Б. Гоулд. – М. : Мир, 1978. – 848 с.
- 3 Хемминг, Р. В. Цифровые фильтры / Р. В. Хемминг. – М. : Сов. радио, 1980. – 224 с.
- 4 Каппелини, В. Цифровые фильтры и их применение / В. Каппелини, А. Константинодис, П. Эмилиани. – М. : Радио и связь, 1983. – 350 с.
- 5 Гольденберг, Л. М. Цифровая обработка сигналов : учеб. пособие для вузов / Л. М. Гольденберг, В. Д. Матюшкин, М. Н. Поляк. – М. : Радио и связь, 1990. – 315 с.
- 6 Сергиенко, А. Б. Цифровая обработка сигналов / А. Б. Сергиенко. – СПб. : Питер, 2002. – 608 с.
- 7 Солонина, А. И. Алгоритмы и процессоры цифровой обработки сигналов / А. И. Солонина, Д. А. Улахович, Л. А. Яковлев. – СПб. : БХВ-Петербург, 2002. – 464 с.
- 8 Оппенгейм, А. Цифровая обработка сигналов / А. Оппенгейм, Р. Шафер. – М. : Техносфера, 2006. – 853 с.
- 9 Лайонс, Р. Цифровая обработка сигналов / Р. Лайонс ; пер. с англ. – 2-е изд. – М. : ООО «Бином-Пресс», 2006. – 656 с.
- 10 Вашкевич, М. И. Косинусно-модулированные банки фильтров с фазовым преобразованием: реализация и применение в слуховых аппаратах / М. И. Вашкевич, И. С. Азаров, А. А. Петровский. – М. : Горячая линия – Телеком, 2014. – 210 с.

### **Цифровая обработка речевых и звуковых сигналов**

- 11 Петровский, А. А. Анализаторы речевых и звуковых сигналов: методы, алгоритмы и практика (с MATLAB-примерами) / А. А. Петровский. – Минск : Бестпринт, 2009. – С. 456.
- 12 Слуховая система / Я. А. Альтман [и др.] ; под общ. ред. Я. А. Альтмана. – Л. : Наука, 1990. – 620 с.
- 13 Азаров, И. С. Мгновенный гармонический анализ. Обработка звуковых и речевых сигналов в системах мультимедиа / И. С. Азаров, А. А. Петровский. – LAP LAMBERT Academic Publishing, 2011. – 163 с.
- 14 Лобанов, Б. М. Компьютерный синтез и клонирование речи / Б. М. Лобанов, Л. И. Цирульник. – Минск : Белорусская наука, 2008. – 316 с.
- 15 Маркел, Дж. Д. Линейное предсказание речи / Дж. Д. Маркел, А. Х. Грэй. – М. : Связь, 1980. – 308 с.
- 16 Марпл, С. Л. Цифровой спектральный анализ и его приложения / С. Л. Марпл / М. : Мир, 1990. – 584 с.

- 17 Abe, T. Sinusoidal model based on instantaneous frequency attractors / T. Abe, M. Honda // *IEEE Transactions on Audio, Speech, and Language Processing*. – 2006. – Vol. 14, №4. – P. 1292–1300.
- 18 Boashash, B. Estimating and interpreting the instantaneous frequency of a signal / B. Boashash // *Proceedings of the IEEE*. – 1992. – Vol. 80, №4. – P. 520–568.
- 19 Bonada, J. Synthesis of the singing voice by performance sampling and spectral models / J. Bonada, X. Serra // *IEEE Signal Processing Magazine*. – March, 2007. – Vol. 24, iss. 2. – P. 67–79.
- 20 Dutoit, T. An Introduction to Text-to-speech Synthesis / T. Dutoit // *The Netherlands : Kluwer Academic Publishers*. – 1997. – 326 p.
- 21 Elhilali, M. A spectro-temporal modulation index (STMI) for assessment of speech intelligibility / M. Elhilali, T. Chi, S. Shamma // *Speech Communication*. – 2003. – Vol. 41, iss. 2–3. – P. 331–348.
- 22 Huang, X. Spoken language processing / X. Huang, A. Acero, H. W. Hon. – New Jersey : Prentice Hall. – 2001. – 1008 p.
- 23 Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation / H. Kawahara [et al.] // *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'08), Las Vegas, USA, March 31–April 4, 2008 / Las Vegas, 2008*. – P. 3933–3936.
- 24 Levine, S. A sines+transients+noise audio representation for data compression and time/pitch scale modifications / S. Levine, J. Smith // *Proceedings of 105-th AES convention on Signal processing, San Francisco, USA, 1998 / San Francisco, 1998*. – №4781. – 21 p.
- 25 Loizou, P. Speech enhancement: theory and practice / P. Loizou. – CRC press Inc., 2007. – 711 p.
- 26 Painter, T. Sinusoidal analysis-synthesis of audio using perceptual criteria / T. Painter, A. Spanias // *EURASIP Journal on Applied Signal Processing*. – 2003. – №1. – P. 15–20.
- 27 Rabiner, L. R. Fundamentals of speech recognition / L. R. Rabiner, B. H. Juang. – New Jersey : Prentice Hall, 1993. – 496 p.
- 28 Serra, X. Musical Sound Modeling with Sinusoids plus Noise / X. Serra // *Musical Signal Processing / Ed. by C. Roads [et al.]*. – Lisse : Swets & Zeitlinger Publishers, 1997. – P. 91–122.
- 29 Stylianou, Y. Continuous probabilistic transform for voice conversion / Y. Stylianou, O. Cappe, E. Moulines // *IEEE Transactions on Speech and Audio Processing*. – 1998. – Vol. 6, №2. – P. 131–142.
- 30 Vaseghi, S. Advanced digital signal processing and noise reduction, second edition / S. Vaseghi // *John Wiley & Sons Ltd, 2000*. – 480 p.
- 31 Zhang, F. Harmonic transform / F. Zhang, G. Bi, Y.Q. Chen // *IEE Proceedings – Vision, Image and Signal Processing*. – Vol.151, №4. – 2004. – P. 257–264.

## **Кодирование изображений на основе трансформационных преобразований**

32 Цифровое сжатие видеоинформации и звука : учеб. пособие / под ред. В. М. Артюшенко. – М. : Издательско-торговая корпорация «Дашков и Ко», 2003. – С. 426.

33 Методы сжатия данных. Устройства архиваторов, сжатие изображений и видео / Д. Ватолин [и др.]. – М. : ДИАЛОГ-МИФИ, 2003. – С. 384.

34 Гонсалес, Р. Цифровая обработка изображений в MATLAB / Р. Гонсалес, Р. Вудс, С. Эддинс. – М. : Техносфера, 2006. – С. 616.

35 Гонсалес, Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М. : Техносфера, 2005. – С.1072.

36 Миано, Дж. Форматы и алгоритмы сжатия изображений в действии / Дж. Миано. – М. : Триумф, 2003 – С. 336.

37 Чобану, М. Многомерные многоскоростные системы обработки сигналов / М. Чобану. – М. : Техносфера, 2009. – С. 480.

38 Loeffler, C. Practical Fast 1-D DCT Algorithms with 11 Multiplications / C. Loeffler, A. Lightenberg, S. Moshutz // IEEE Conference on Acoustics, Speech, and Signal Processing. – Glasgow, 1989. – Vol. 2. – P. 988–991.

39 Rao, K. Techniques and Standards for Image, Video, and Audio Coding / K. Rao, J. Hwang. – Upper Saddle Rive : Prentice Hall, 1996. – P. 563.

40 Suzuki, T. Integer DCT Based on Direct-Lifting of DCT-IDCT for Lossless-to-Lossy Image Coding / T. Suzuki, M. Ikehara // IEEE Transactions on image processing. – November, 2010. – Vol. 19, №11. – P. 2958–2965.

41 Sweldens, W. The lifting scheme: A custom-design construction of biorthogonal wavelets / W. Sweldens // Applied Computational Harmonic Analysis. – April, 1996. – Vol. 3, №2. – P. 186–200.

42 Vaidyanathan, P. P. Multirate systems and filter banks / P. P. Vaidyanathan. – Prentice-Hall, Englewood Cliffs, NJ, USA. – 1993. – 928 p.

43 Kovacevic, J. Wavelet and Subband Coding / J. Kovacevic, M. Vetterli // Prentice-Hall signal processing series. – Englewood Cliffs, NJ: Prentice Hall PTR, 1995. – P. 488.

44 Парфенюк, М. Параунитарные банки фильтров на основе алгебры кватернионов: теория и применение / М. Парфенюк, А. А. Петровский // Цифровая обработка сигналов. – 2008. – №1. – С. 22–36.

45 Parfieniuk, M. Inherently lossless structures for eight- and sixchannel linear-phase paraunitary filter banks based on quaternion multipliers / M. Parfieniuk, A. Petrovsky // Signal Process. – 2010. – Vol. 90. – P. 1755–1767.

46 Suzuki, T. Lifting-based paraunitary filterbanks for lossy/lossless image coding / T. Suzuki, Y. Tanaka, M. Ikehara // Proc. EUSIPCO. – 2006. – Vol. 6.