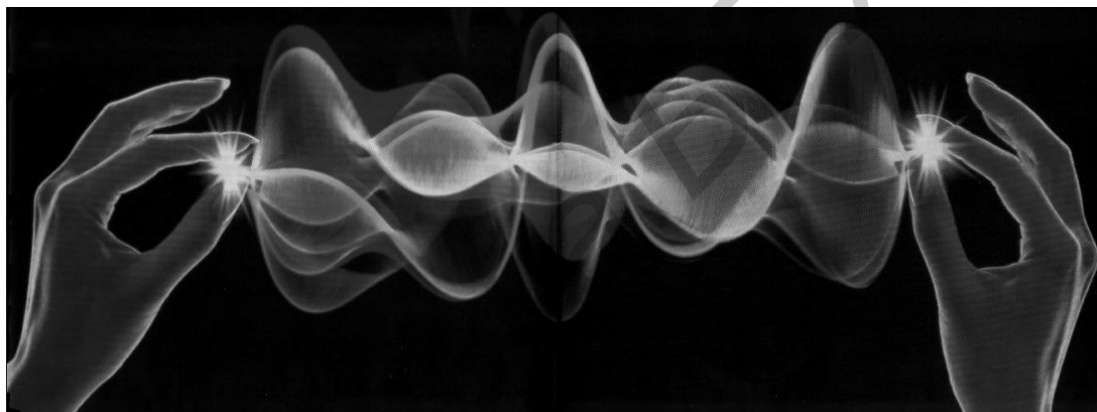


Министерство образования Республики Беларусь  
Учреждение образования  
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ»

Кафедра электронных вычислительных средств

*Цифровые банки фильтров: анализ, синтез и применение в  
мультимедиа системах*



**Учебно-методическое пособие**

по курсам «Теория и применение цифровой обработки сигналов»,  
«Речевые интерфейсы ЭВС»

для студентов специальности I – 40 02 02  
«Электронные вычислительные средства»  
дневной формы обучения

Минск 2006

УДК 621.391.2 (075.8)

ББК 32.88 я 73

Ц 75

Авторы:

А.А. Петровский, М.Парфенюк, А.Борович, М.З. Лившиц

Рецензенты:

зав. кафедрой ИИТ БГУИР, д.т.н. В.В. Голенков

зав. кафедрой ЭВМ БГУИР, д.т.н. Р.Х. Садыхов

**Цифровые** банки фильтров: анализ, синтез и применение в мультимедиасистемах: Учеб. метод. пособие для студ. спец. I–40 02 02 «Электронные вычислительные средства» дневной формы обуч. / А.А. Петровский, М.Парфенюк, А.Борович и др. – Мн.: БГУИР, 2006. – 83с.: ил.  
ISBN 985-488-017-6

Описываются эффективные структуры и методы проектирования банков цифровых фильтров. Подробно рассматриваются ДПФ-модулированные и косинусно-модулированные банки фильтров как равнополосные, так и неравнополосные на основе фазовых звеньев. Рассматривается ДПФ с неравномерным частотным разрешением (Warped Discrete-Fourier Transform – WDFT), позволяющее аппроксимировать психоакустические шкалы для построения перцептуально мотивированных систем сжатия и передачи речевой информации, а также систем повышения качества речи. Все примеры сопровождаются программной реализацией в среде MATLAB. Особое внимание уделено практическому применению банков цифровых фильтров и WDFT в реальных приложениях систем мультимедиа: компрессии речевых сигналов и редактирования шума – как препроцессоров систем кодирования речи.

УДК 621.391.2 (075.8)

ББК 32.88 я 73

ISBN 985-488-017-6

© Коллектив авторов, 2006

© БГУИР, 2006

## Содержание

<b>1. БАНКИ ЦИФРОВЫХ ФИЛЬТРОВ</b> .....	<b>5</b>
1.1. Определения .....	5
1.2. ДПФ-модулированные банки фильтров .....	12
1.2.1. ДПФ-модулированный равнополосный банк фильтров на основе полифазной структуры фильтров дециматоров и интерполяторов.....	12
1.2.2. ДПФ-модулированный неравнополосный банк фильтров.....	17
1.2.3. Реализация ДПФ-модулированных банков фильтров в среде MATLAB.....	23
1.2.4. Пример банка ДПФ-модулированных полифазных фильтров.....	28
1.3. Косинусно-модулированные банки фильтров.....	30
1.3.1. Проектирование косинусно-модулированного банка фильтров.....	30
1.3.2. Косинусно-модулированные банки фильтров с квазиполным восстановлением.....	32
1.3.3. Реализация косинусно-модулированных банков фильтров в среде MATLAB .....	35
<b>2. ЭФФЕКТИВНОСТЬ СУБПОЛОСНОГО КОДИРОВАНИЯ СИГНАЛА</b> .....	<b>39</b>
2.1. Дисперсия ошибки реконструкции сигнала .....	39
2.2. Обобщенный банк фильтров .....	40
2.3. Ортогональный банк фильтров .....	42
2.3.1. Дисперсия ошибки.....	42
2.4. Оптимальное распределение бит в каналах банка фильтров.....	43
2.4.1. Метод множителей Лагранжа .....	43
2.4.2. Оптимальное распределение бит .....	46
2.5. Оценка эффективности субполосного кодирования .....	48
2.5.1. Сигнал с нормальным законом распределения.....	49
2.5.2. Банк фильтров с равными коэффициентами децимации.....	49
2.5.3. Ортогональный банк фильтров.....	49
2.5.4. Входной сигнал с нормальным законом распределения и ортогональный банк фильтров с равными коэффициентами децимации .....	50
<b>3. ДПФ С НЕРАВНОМЕРНЫМ ЧАСТОТНЫМ РАЗРЕШЕНИЕМ (WDFT)</b> .....	<b>51</b>
3.1.1. Основные принципы WDFT.....	51
3.1.2. Обратное преобразование WDFT .....	54
3.1.3. Оценка ошибки реконструкции WDFT-синтеза .....	56
3.1.4. Формирование избыточного базиса WDFT .....	58
3.1.5. Аппроксимация психоакустических шкал.....	60

<b>4. ПРИМЕНЕНИЕ БАНКОВ ЦИФРОВЫХ ФИЛЬТРОВ И WDFT В СИСТЕМАХ МУЛЬТИМЕДИА.....</b>	<b>64</b>
4.1. Перцептуально мотивированный широкополосный кодер речи.....	64
4.1.1. Структура кодера и декодера.....	64
4.1.2. Разбиение частотной оси на субполосы.....	66
4.1.3. Реконфигурирование структуры кодовой книги на основе субполосной перцептуальной энтропии .....	67
4.2. Система редактирования шума в речевом сигнале.....	71
4.2.1. Введение .....	71
4.2.2. Краткое описание системы .....	72
4.2.3. Психоакустически мотивированное правило спектрального взвешивания.....	73
4.2.4. Оценка СПМ шума.....	74
4.2.5. Оценка качества системы подавления шума .....	76
4.3. Комбинированная система кодирования широкополосной речевой информации .....	77
<b>ЛИТЕРАТУРА .....</b>	<b>80</b>

# 1. Банки цифровых фильтров

## 1.1. Определения

**Фильтр** (от англ. – «Filter») – линейная стационарная система, то есть свойства фильтра не зависят от времени. Независимость свойств фильтра от времени означает, что задержка входа приводит к такой же задержке выхода. К основным характеристикам фильтров относятся: импульсная характеристика  $h(n)$ , передаточная характеристика  $H(z) = \sum_n h(n)z^{-n}$ , частотная характеристика  $H(e^{j\omega}) = \sum_n h(n)e^{-j\omega n}$ , порядок фильтра.

**Банк фильтров (БФ)** (от англ. – «Filter Bank») – цифровое устройство, в котором входной сигнал, представленный последовательностью отсчетов, с помощью  $M$  различных цифровых фильтров разбивается на  $M$  различных канальных сигналов (для обработки некоторым способом каждого из них), из которых с помощью выходных фильтров и последующего суммирования образуется последовательность отсчетов выходного сигнала.

Основную идею построения системы анализа/синтеза сигналов с использованием банка фильтров раскрывает рис. 1.1.

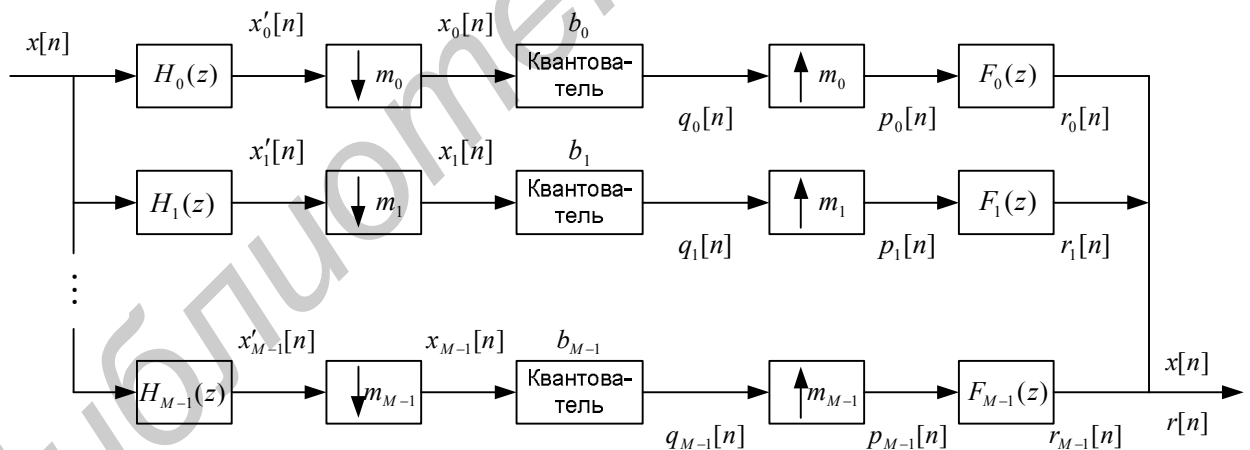


Рис. 1.1. Система анализа/синтеза сигналов на основе банка фильтров

Исходный сигнал разбивается при помощи фильтров анализа  $H_k(z)$ ,  $k = 0, 1, \dots, M-1$  на  $M$  субполосных составляющих, которые в идеальном случае в частотной области не перекрываются. Подобранным соответствующим образом набор фильтров синтеза  $F_k(z)$ ,  $k = 0, 1, \dots, M-1$ , можно восстановить исходный сигнал из его субполосных компонент. Вследствие ограничения ширины спектра сигналов на выходе БФ можно уменьшить частоту дискретизации субполосных сигналов пропорционально уменьшению ширины спектра. Для

понижения частоты дискретизации на стадии анализа и последующего повышения на стадии синтеза используются соответственно компрессоры и экспандеры частоты дискретизации. В случае, когда коэффициент прореживания в каждом канале равен отношению ширины спектра субполосного сигнала к ширине спектра исходного или

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

говорят о системе анализа/синтеза с полной децимацией. Таким образом, получается  $M$  сигналов, отражающих поведение исходного сигнала в каждом частотном поддиапазоне, которые представлены в сумме тем же количеством отсчетов, что и исходный сигнал. Каждый субполосный сигнал в отдельности может быть эффективно обработан по некоторому алгоритму  $b_k$ ,  $k = 0, 1, \dots, M - 1$ .

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

Равнополосная декомпозиция подразумевает одинаковый коэффициент децимации и одинаковую суммарную ширину полосы пропускания каждого канала. В этот класс цифровых БФ входят также банки с многокомпонентными фильтрами (имеющими более одной полосы пропускания). В случае неравнополосных каналов коэффициенты децимации различны и в общем случае могут быть выражены в виде рационального числа

$$m_k = \frac{q_k}{p_k}. \quad (1.2)$$

В связи с тем, что невозможно построить фильтр с идеальной прямоугольной АЧХ, приходится мириться с наличием у полосового фильтра переходной зоны и неравномерностью в полосе пропускания. Это обстоятельство в сочетании с эффектом «заворачивания» частот, возникающим при децимации, может привести к значительным искажениям при восстановлении сигнала. Однако существуют методики расчета БФ, которые позволяют полностью (БФ с полным восстановлением) или почти полностью, то есть с любой наперед заданной точностью, восстановить исходный сигнал.

**Дециматор (компрессор частоты дискретизации)** – устройство, осуществляющее децимацию (прореживание) сигнала во времени. Децимация

(от англ. – «Decimation») – операция, заключающаяся в исключении (пропуске) отсчетов входного сигнала, с порядковым номером, кратным коэффициенту децимации. Децимация в  $M$  раз обозначается обычно как  $(\downarrow M)$ :  $y(n) = (\downarrow M)x(n) = x(nM)$ . В частотной области это запишется как  $Y(e^{j\omega}) = \frac{1}{M} \sum_{k=0}^{M-1} X\left(e^{j(\omega-2\pi k)}\right)$ , то есть спектр выходного сигнала операции децимации содержит  $M$  копий «расширенного» в  $M$  раз спектра входного сигнала, как это показано на рис. 1.2.

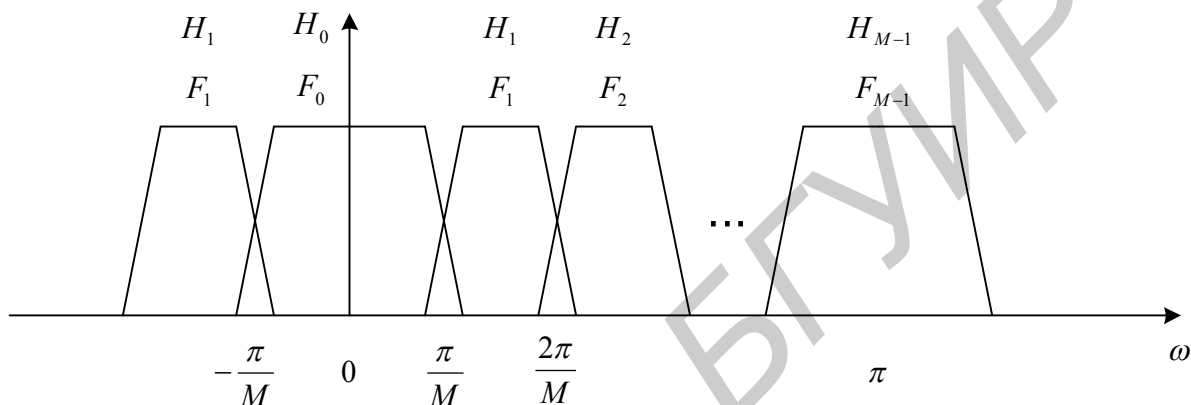


Рис. 1.2. Децимация сигнала в  $M$  раз

Как видно из рис. 1.2, если сигнал не ограничен полосой частот  $-\frac{\pi}{M} < \omega < \frac{\pi}{M}$ , то происходит наложение спектров копий, то есть *алайзинг* (от англ. – «aliasing»). Поэтому в банке фильтров перед децимацией выполняется НЧ-фильтрация. Совокупность фильтра и дециматора называется *фильтром-дециматором*.

**Интерполятор** – устройство, выполняющее действия, обратные децимации. Интерполяция (от англ. – «Interpolation») – операция, заключающаяся во встраивании (добавлении) между отсчетами, чей порядковый номер кратен определенному числу, некоторой константы (обычно нуля). Интерполяция в  $M$  раз обычно обозначается как  $(\uparrow M)$ :

$$u(n) = \begin{cases} \text{если } n/M, \text{ тогда } x(n/M) \\ \text{иначе } 0 \end{cases} . \text{ В частотной области это запишется как}$$

$U(e^{j\omega}) = X(e^{jM\omega})$ , то есть спектр выходного сигнала операции интерполяции содержит  $M$  копий «сжатого» в  $M$  раз спектра входного сигнала. Эти копии повторяются через  $\omega_k = \frac{2\pi k}{M}$ . Для их устранения после интерполятора ставится

НЧ-фильтр. Совокупность интерполятора и фильтра называется **фильтром-интерполятором**.

**Эквивалентные схемы включения интерполяторов и дециматоров.**

Два «замечательных» тождества (от англ. – «Noble Identity») – позволяют уменьшить число вычислений в банке фильтров. Это обусловлено тем, что частота дискретизации входного сигнала снижается при помощи дециматора, а затем осуществляется процесс фильтрации, т.о. общая вычислительная сложность уменьшается пропорционально коэффициенту децимации. После окончания обработки субполосного сигнала в каждом канале банка фильтров частота дискретизации повышается при помощи интерполятора. Данные равенства (эквивалентные схемы включения) широко применяются для вывода различных соотношений и представлены на рис. 1.3.

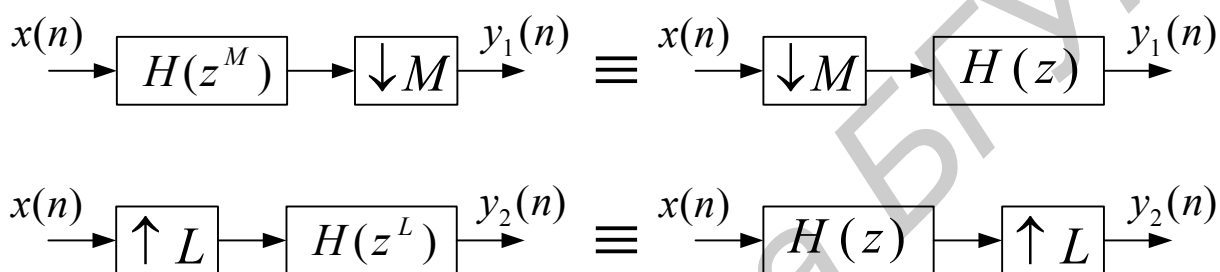


Рис. 1.3. Эквивалентные схемы включения фильтров-дециматоров и фильтров-интерполяторов

**Приведение коэффициентов децимации.**

Если банк фильтров имеет неравные коэффициенты децимации ( $m_k$  не равны как на рис. 1.1), он может быть трансформирован в банк фильтров с равными коэффициентами децимации (все  $m_k$  равны) как описано ниже.

Пусть  $N = \text{lcm}(m_0, m_1, \dots, m_{M-1})$ . Канал с коэффициентом децимации/интерполяции  $m_k$  показан на рис. 1.4. Представим систему с цепочкой задержек с коэффициентом передискретизации  $\frac{N}{m_k}$  между стороной анализа и синтеза, как показано на рис. 1.5.

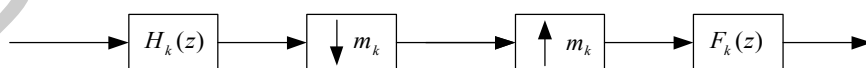


Рис. 1.4. К-й канал банка фильтров



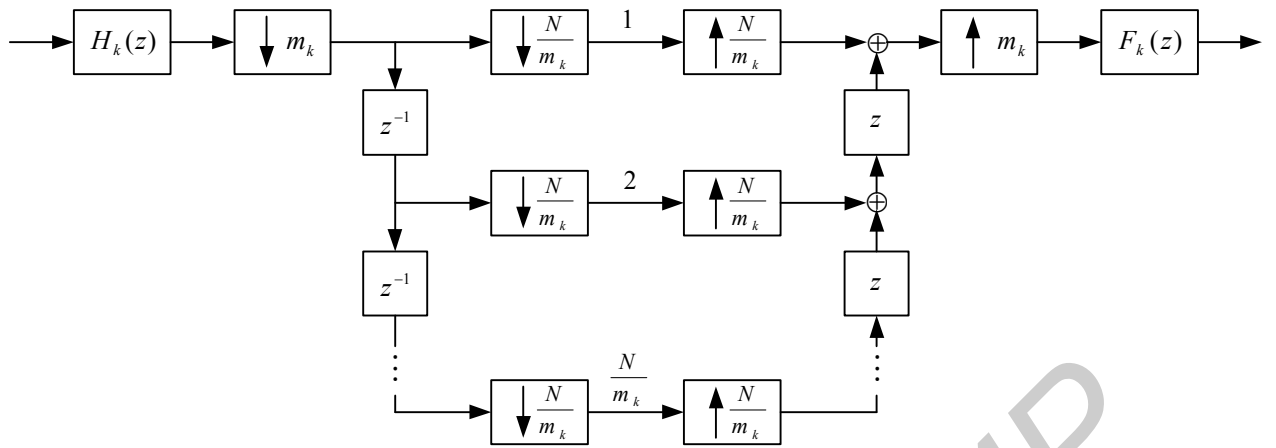


Рис. 1.5. К-й канал с цепочкой задержек

Обе системы, представленные на рис. 1.4 и рис. 1.5 идентичны, так как система с цепочкой задержек обладает совершенной реконструкцией и имеет нулевую задержку (положительная степень  $z$  используется на синтезирующей стороне). Перестановка задержек в представленной системе до  $m_k$ -дециматора и после  $m_k$ -интерполятора приводит к системе, представленной на рис. 1.6.

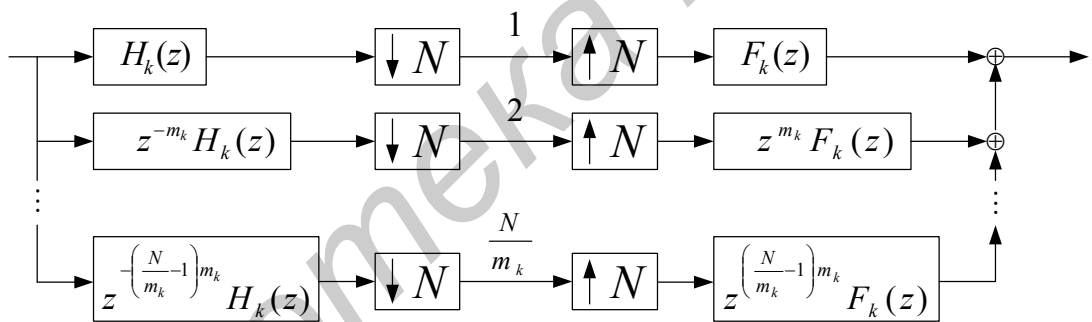


Рис. 1.6. Эквивалентная система к-го канала

Фактически один канал с коэффициентом децимации  $m_k$  заменяется на  $\frac{N}{m_k}$  каналов с коэффициентов децимации  $N$ . Подобным способом все каналы  $0 \leq k \leq M-1$  могут быть преобразованы к системам с коэффициентом децимации  $N$ . Используя уравнение (1.1), максимальное количество каналов можно представить следующим выражением

$$\frac{N}{m_0} + \frac{N}{m_1} + \dots + \frac{N}{m_{M-1}} = N. \quad (1.3)$$

Таким образом, банк фильтров с неравными коэффициентами децимации выражается как банк фильтров с равными коэффициентами децимации, имеющий  $N$  каналов.

**Полное восстановление.** Полное восстановление или перфективная реконструкция (от англ. – «Perfect Reconstruction») – свойство банка цифровых фильтров, заключающееся в том, что сигнал, прошедший через схему анализа-синтеза идентичен входному с точностью до задержки. Для этого фильтры синтеза должны подавлять алаизинг и устранять амплитудные и фазовые искажения:

$$\begin{aligned} F_0(z)H_0(z) + F_1(z)H_1(z) &= 2z^{-l}, \quad l \bmod 2 \neq 0, \\ F_0(z)H_0(-z) + F_1(z)H_1(-z) &= 0, \end{aligned} \quad (1.4)$$

где  $F_k$  - синтезирующий фильтр  $k$ -го канала,  $H_k$  - анализирующий фильтр  $k$ -го канала.

Модуляционная матрица должна удовлетворять соотношению

$$F_m(z)H_m(z) = \begin{bmatrix} z^{-l} & 0 \\ 0 & z^{-l} \end{bmatrix}. \quad (1.5)$$

### Понятие параунитарной системы.

Банк фильтров является **ортогональным (параунитарным)**, когда анализирующие и синтезирующие фильтры и их соответственно смещенные версии ортогональны друг другу.

Пусть  $x[n] = \begin{bmatrix} x_0[n] \\ \vdots \\ x_{N-1}[n] \end{bmatrix}$  будет входным вектором и  $y[n] = \begin{bmatrix} y_0[n] \\ \vdots \\ y_{N-1}[n] \end{bmatrix}$  будет

соответствующим выходным вектором с  $N \times N$  параунитарной передаточной матрицей  $A(z)$ . Пусть  $S_{xx}(e^{jw})$  будет  $N \times N$  СПМ-матрица входного вектора  $x[n]$ . Заметим, что СПМ  $i$ -го входного компонента  $x_i[n]$  является  $i$ -й элемент  $S_{xx}(e^{jw})$ . Поэтому дисперсия  $x_i[n]$  составит

$$\int_{-\pi}^{\pi} (S_{xx}(e^{jw}))_i \frac{dw}{2\pi}. \quad (1.6)$$

Усредненная дисперсия входного сигнала будет

$$\frac{1}{N} \sum_{i=0}^{N-1} \sigma_i^2 = \frac{1}{N} \int_{-\pi}^{\pi} \text{tr}(S_{xx}(e^{jw})) \frac{dw}{2\pi}, \quad (1.7)$$

где  $\text{tr}(B)$  – след матрицы  $B$ .

СПМ вектора на выходе определяется как

$$S_{yy}(e^{jw}) = A(e^{jw})S_{xx}(e^{jw})A^H(e^{jw}), \quad (1.8)$$

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

Усредненная дисперсия выхода составит

$$\frac{1}{N} \int_{-\pi}^{\pi} \text{tr}(S_{yy}(e^{jw})) \frac{dw}{2\pi} = \frac{1}{N} \int_{-\pi}^{\pi} \text{tr}(A(e^{jw})S_{xx}(e^{jw})A^H(e^{jw})) \frac{dw}{2\pi}. \quad (1.9)$$

Так как  $\text{tr}(AB) = \text{tr}(BA)$ , это упрощает выражение

$$\frac{1}{N} \int_{-\pi}^{\pi} \text{tr}(S_{xx}(e^{jw})A^H(e^{jw})A(e^{jw})) \frac{dw}{2\pi}. \quad (1.10)$$

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

$$\frac{1}{N} \int_{-\pi}^{\pi} \text{tr}(S_{xx}(e^{jw})) \frac{dw}{2\pi} \quad (1.11)$$

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

Известно, что для ортогонального банка фильтров с равными коэффициентами децимации полифазные матрицы анализа и синтеза являются параунитарными. Полифазная матрица, определенная для банков фильтров с равным коэффициентом децимации, может быть расширена до неравных коэффициентов децимации (рис. 1.1), так как для схемы на рис. 1.6 может быть сконструирована ее эквивалентная  $N \times N$  матрица. Эти полифазные матрицы также параунитарные. Подобный ортогональный банк с неравными коэффициентами децимации может быть построен, используя несколько ортогональных банков фильтров с равными коэффициентами децимации в древовидной структуре.

## 1.2. ДПФ-модулированные банки фильтров

### 1.2.1. ДПФ-модулированный равнополосный банк фильтров на основе полифазной структуры фильтров дециматоров и интерполяторов

При проектировании банка цифровых фильтров с равнополосными каналами используется модель с полной модуляцией. Если банк основан на одном НЧ-фильтре-прототипе с конечной импульсной характеристикой (КИХ), то ширина полосы пропускания фильтра-прототипа определяет ширину каждого канала. Чтобы такой фильтр мог выделить полосу, соответствующую каждому каналу, необходимо сдвинуть спектр в область низких частот при помощи гетеродина (экспоненциального модулятора), а затем осуществить НЧ-фильтрацию фильтром-прототипом. После чего можно снизить частоту дискретизации субполосного сигнала без потери информации. Снижение частоты дискретизации осуществляет компрессор посредством децимации, который удаляет  $M - 1$  отсчетов из каждой последовательности длиной  $M$ . Максимальный коэффициент децимации равен количеству каналов  $K$ , таким образом, данный банк фильтров является максимально децимированным.

Синтез осуществляется в обратной последовательности. Сначала увеличивается частота дискретизации. В экспандере между каждыми двумя отсчетами вставляются  $M - 1$  нулевых отсчетов. Затем осуществляется фильтрация субполосных сигналов с последующей модуляцией с целью перемещения субполосы в соответствующий частотный диапазон, который она занимала в исходном широкополосном сигнале. Суммирование выходов всех каналов синтезирующего банка фильтров дает восстановленный широкополосный сигнал (рис. 1.7).

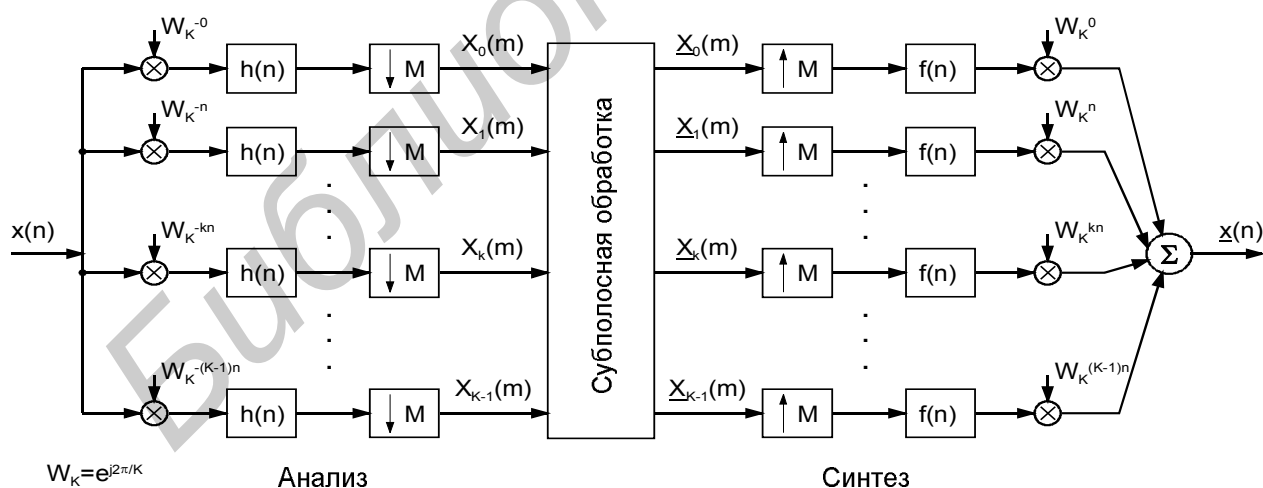


Рис. 1.7. Модель ДПФ-модулированного банка фильтров, основанного на полной модуляции

НЧ-фильтр-прототип может быть спроектирован стандартными методами, такими как синтез при помощи взвешивающих окон, частотной выборки и т.д. Частота среза фильтра-прототипа определяет количество каналов и их ширину, так как вся полоса может быть поделена на  $K$  равных частей. Степень наложения субполос ограничивается в соответствии с требованиями, налагаемыми областью применения конкретного банка фильтров.

Импульсная и частотная характеристика фильтра для каждого канала определяется следующим образом:

$$h_k(n) = h(n)W_K^{kn} = h(n)e^{j2\pi kn/K}, \quad (1.12)$$

$$H_k(e^{j\omega}) = H\left(e^{j\left(\omega - \frac{2\pi k}{K}\right)}\right). \quad (1.13)$$

Система характеризуется равномерным размещением полос с шагом

$$\omega_k = \frac{2\pi k}{K} \quad k = 0, 1, \dots, K-1. \quad (1.14)$$

Непосредственная реализация такой схемы банка фильтров является крайне неэффективной. В каждом канале при большой частоте дискретизации осуществляется свертка с импульсной характеристикой фильтра-прототипа, что приводит к значительному увеличению вычислительных затрат, которые можно снизить путем снижения частоты дискретизации.

Ключом для построения эффективной структуры является полифазная декомпозиция фильтра-прототипа. Она основывается на разбиении, децимации, группировании коэффициентов фильтра на подгруппы, называемые **полифазными фильтрами** (рис. 1.8).

$$H(z) = \sum_{k=0}^{M-1} z^{-k} G_k(z^M) = G_0(z^M) + z^{-1}G_1(z^M) + \dots + z^{-(M-1)}G_{M-1}(z^M) \quad (1.15)$$

$$G_k(z) = h(k) + h(k+M)z^{-1} + h(k+2M)z^{-2} + \dots$$

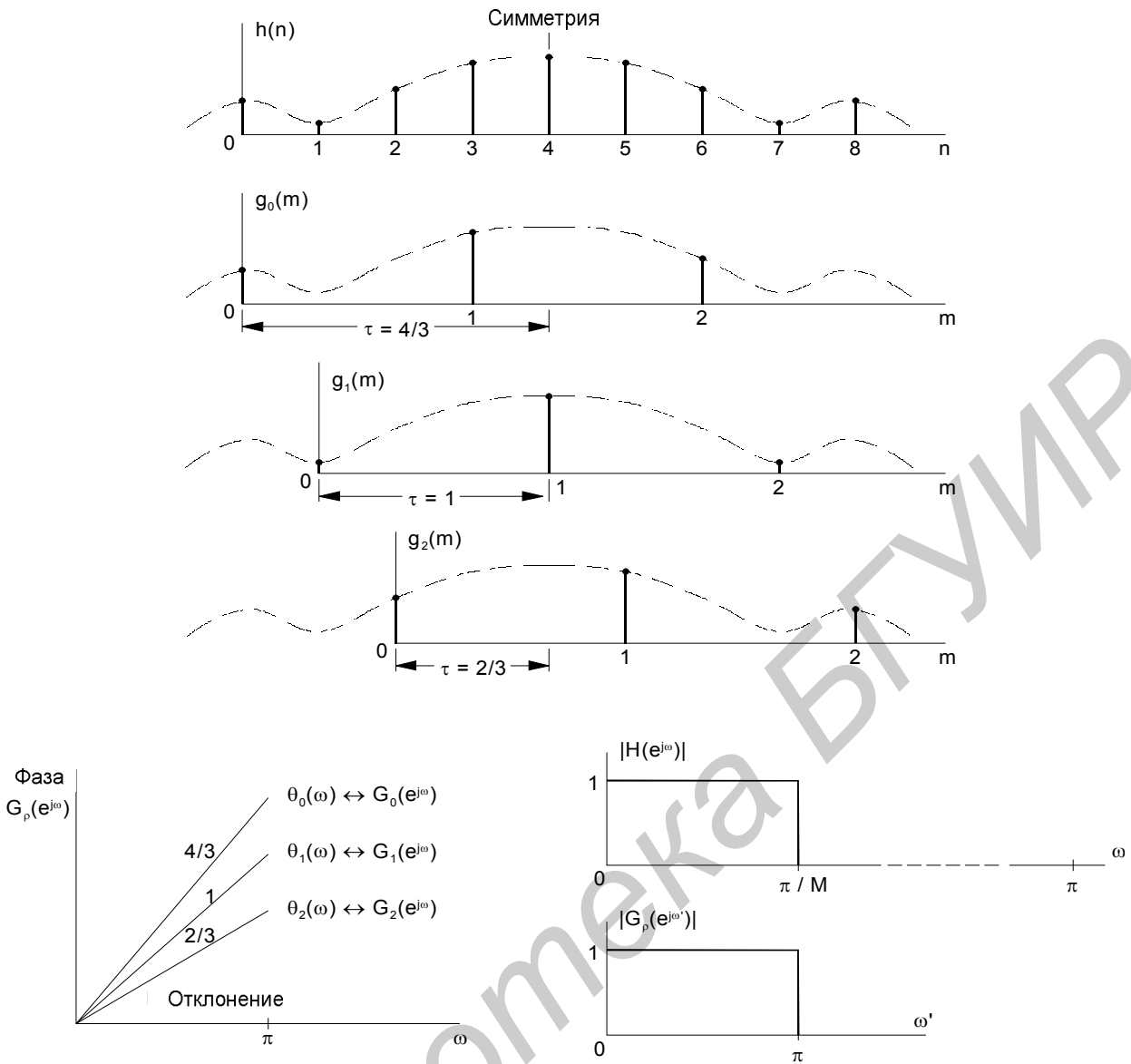


Рис. 1.8. Полифазная декомпозиция фильтра-прототипа

Такое группирование полифазных фильтров может быть поделено между каналами (рис. 1.9 и 1.10).

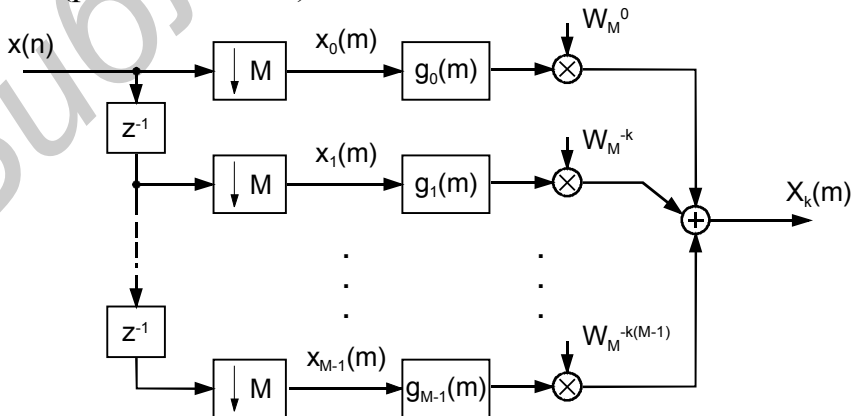


Рис. 1.9. Полифазная структура канала максимально децимированного банка анализа

Оценка субполосных сигналов осуществляется после модуляции, реализуемой при помощи ДПФ, вычислительную сложность которого можно ограничить, используя алгоритмы БПФ.

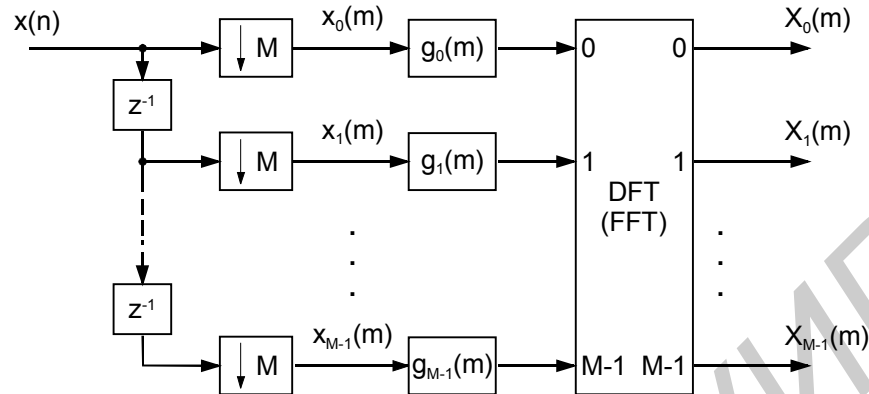


Рис. 1.10. ДПФ-модулированный банк фильтров анализа с равнополосными каналами на основе полифазной структуры фильтров дециматоров

Полученная структура является весьма эффективной. Ее сложность можно оценить по следующей формуле:

$$Z = M \frac{L}{M} + M \log_2 M = L + M \log_2 M \quad [\text{операций/входной отсчет}]. \quad (1.16)$$

Сложность прямой реализации в случае максимальной децимации составит

$$Z' = KM(L+1) = M^2(L+1) \Big|_{K=M} \quad [\text{операций/входной отсчет}]. \quad (1.17)$$

Сравнение вычислительной сложности для прямой и модифицированной структуры банка фильтров представлено в табл. 1.1 и на рис. 1.11 и рис. 1.12.

Таблица 1.1.

Сравнение вычислительной сложности для прямой и модифицированной структуры банка фильтров

$Z'/Z$		$L$ (длина импульсной характеристики фильтра-прототипа)			
		64	128	256	512
$M$ (кол-во каналов)	4	14.4444	15.1765	15.5758	15.7846
	8	47.2727	54.3158	58.7429	61.2537
	16	130.0000	172.0000	205.6000	228.0000
	32	297.1429	458.6667	632.6154	781.7143

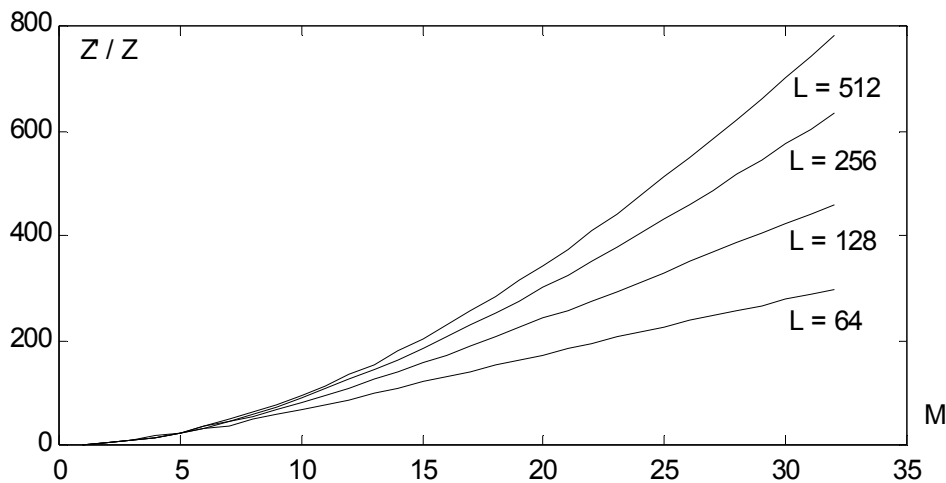


Рис. 1.11. Зависимость вычислительной сложности от количества каналов

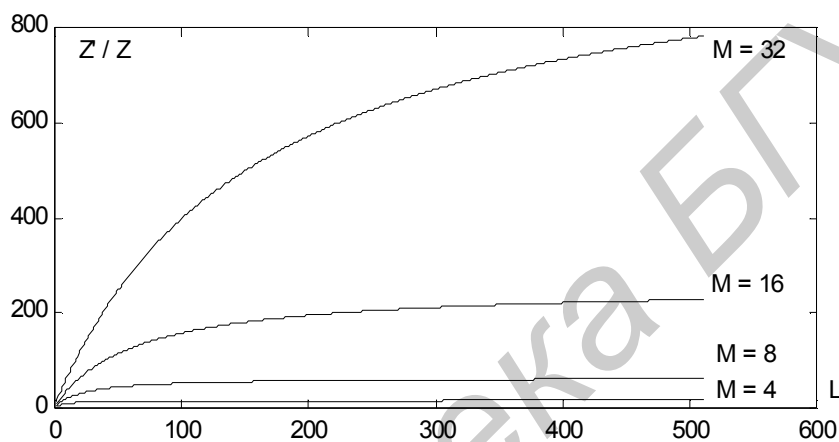


Рис. 1.12. Зависимость вычислительной сложности от длины импульсной характеристики фильтра-прототипа

ДПФ-модулированный банк фильтров синтеза и АЧХ представлены на рис. 1.13 и рис. 1.14, соответственно.

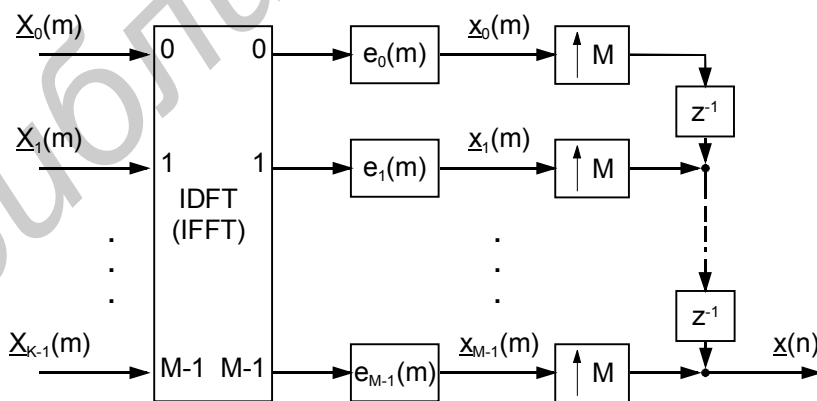


Рис. 1.13. ДПФ-модулированный банк фильтров синтеза с равнополосными каналами на основе полифазной структуры фильтров дециматоров



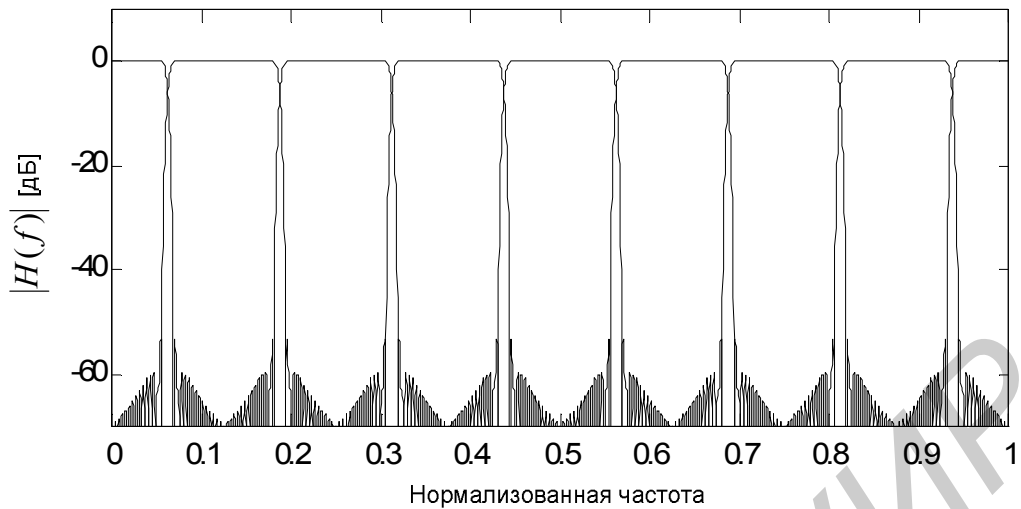


Рис. 1.14. АЧХ 8-канального ДПФ-модулированного банка цифровых фильтров с равнополосными каналами

### 1.2.2. ДПФ-модулированный неравнополосный банк фильтров

В ряде случаев разбиение исходной полосы сигнала на субполосы равной ширины не всегда является необходимым. Примером этого может служить перцептуальное преобразование, где система должна отражать неравномерное разложение на критические частотные субполосы согласно особенностям человеческого восприятия акустической информации.

Описанный выше ДПФ-модулированный банк фильтров с равнополосными каналами можно преобразовать в банк с неравнополосными каналами разложением субполос при помощи фазового фильтра (звена) (от англ. – «All-pass filter»):

$$z^{-1} \Rightarrow A(z) = H_{AP}(z) = \frac{z^{-1} - a}{1 - az^{-1}}. \quad (1.18)$$

Каскадное расположение фазовых звеньев становится причиной смещения частотных компонент сигнала. Фазовое звено не влияет на амплитуды частотных компонент сигнала, а лишь изменяет их фазу, при этом фазовая характеристика данного звена является нелинейной функцией частоты:

$$\left| A(e^{j\omega}) \right| = \left| e^{j\varphi(\omega)} \right| = 1, \quad (1.19)$$

$$\varphi(\omega) = -\omega - 2 \arctan \frac{|a| \sin(\omega - a)}{1 - |a| \cos(\omega - a)}. \quad (1.20)$$

Характер нелинейности фазовой характеристики определяется параметром фильтра  $a$ . Фильтр является стабильным, при  $|a| < 1$ . Чем ближе значение модуля параметра к 1, тем сильнее выражена нелинейность фазовой характеристики фильтра. Для параметра  $a = 0$  задержка одинакова для всех частотных составляющих сигнала. Компоненты входного сигнала с разными частотами, прошедшие через фазовый фильтр, задерживаются неодинаково, так как групповая задержка фильтра также является нелинейной функцией от пульсации:

$$\tau(\omega) = -\frac{d\varphi(\omega)}{d\omega} = \frac{-(1 - |a|^2)}{(1 - |a|^2) + 2|a|(1 - \cos(\omega - a))}. \quad (1.21)$$

Последовательность, прошедшая через ряд таких фильтров не является оригинальной задержанной последовательностью. Компоненты сигнала, зависящие от своей пульсации, перемещаются медленнее или быстрее. Таким образом, сигнал, прошедший через последовательность фазовых звеньев, является деформированной в частотной области версией исходного входного сигнала. Еще более гибкого разложения на субполосы можно достичь, применяя фазовые фильтры высших порядков  $R > 1$ :

$$H_{APR}(z) = e^{ja_0} \prod_{r=1}^R \frac{z^{-1} - a_r^*}{1 - \underline{a}_r z^{-1}}. \quad (1.22)$$

В таких фильтрах фазовая функция является более эластичной, что позволяет осуществлять более свободную деформацию субполос в банке фильтров (рис. 1.15 – 1.20).

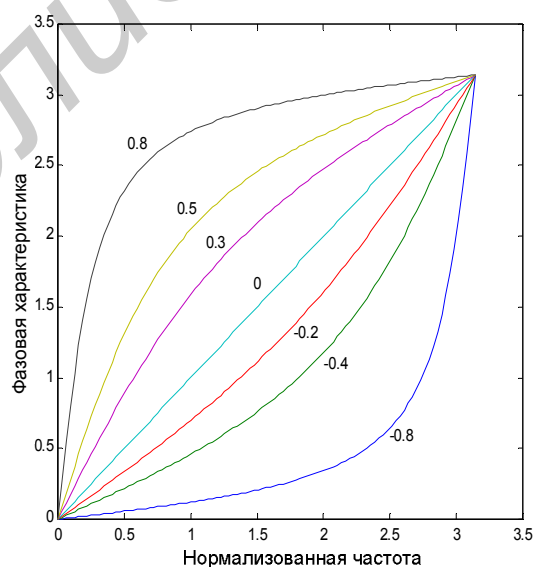


Рис. 1.15. Фазовая характеристика фазового звена первого порядка для различных коэффициентов фильтра

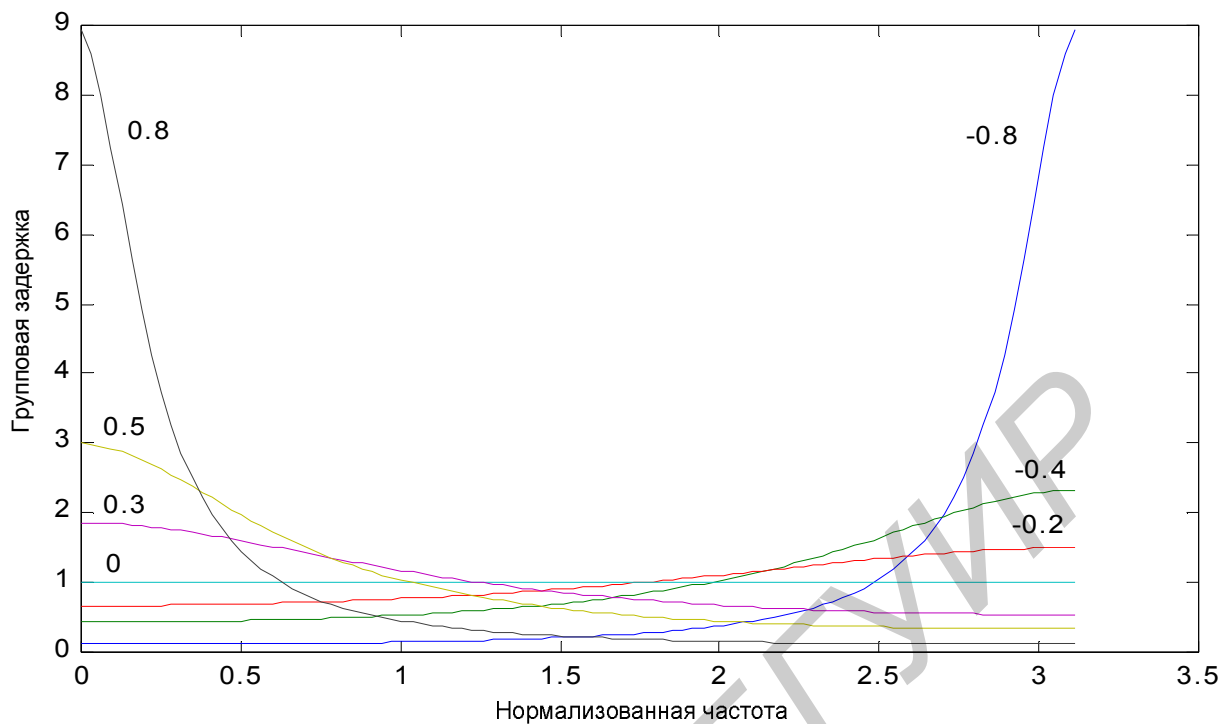


Рис. 1.16. Групповая задержка фазового фильтра для различных значений коэффициентов

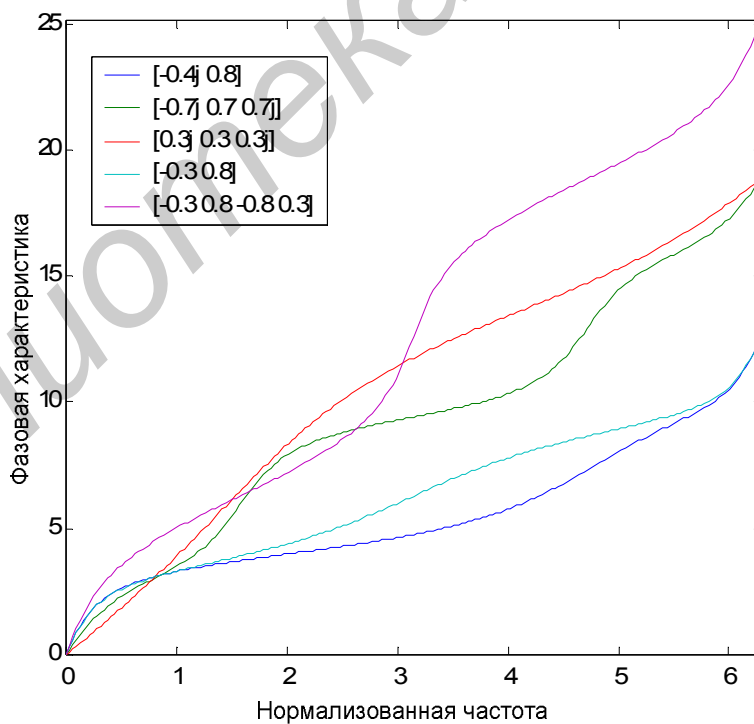


Рис. 1.17. Фазовая характеристика фазовых фильтров высших порядков для различных коэффициентов при многократном перекрытии

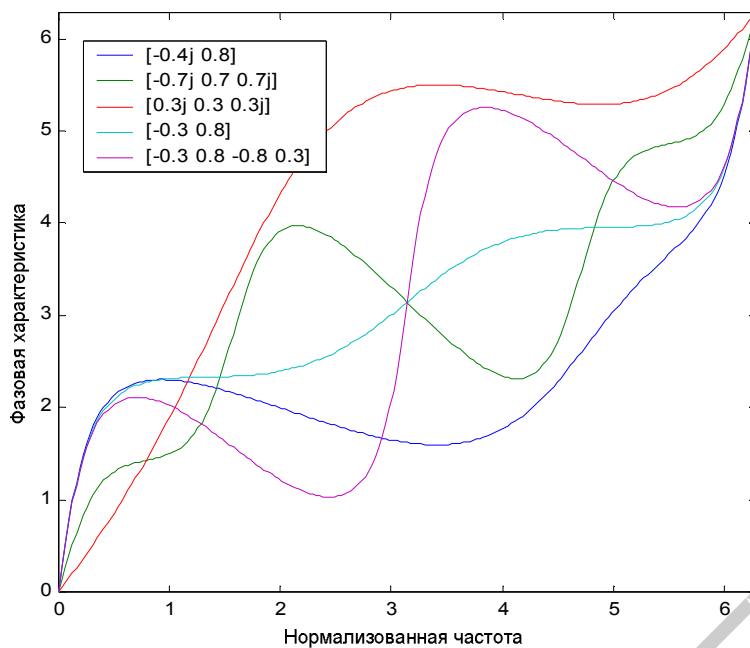


Рис. 1.18. Фазовая характеристика фазовых фильтров высших порядков для различных коэффициентов при однократном перекрытии

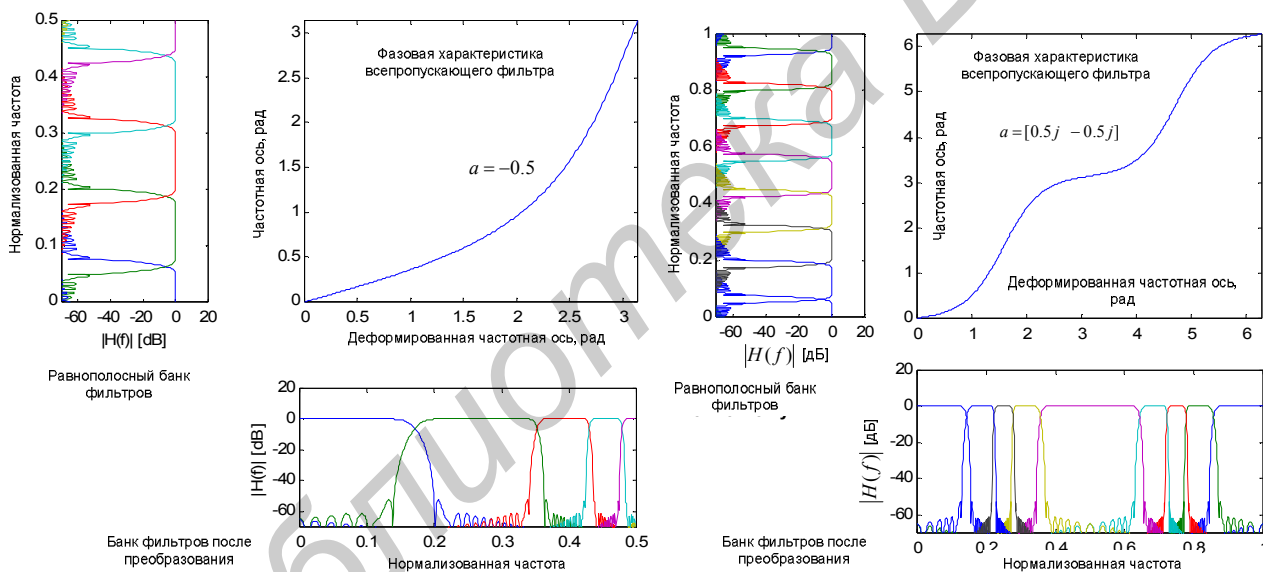


Рис. 1.19. Формирование субполос в банке фильтров с фазовыми звеньями

Пример структуры ДПФ-модулированного неравнополосного банка фильтров анализа-синтеза представлен на рис. 1.21.

В приложениях кодирования речевых сигналов ДПФ-модулированные банки не применяются. Чаще используются банки с косинусной модуляцией. Применяемая в них модуляция реализуется косинусоидальными функциями, в результате чего входной речевой сигнал после анализа остается действительным (сохраняется его структура). Принцип модуляции в банках цифровых фильтров представлен на рис. 1.22.

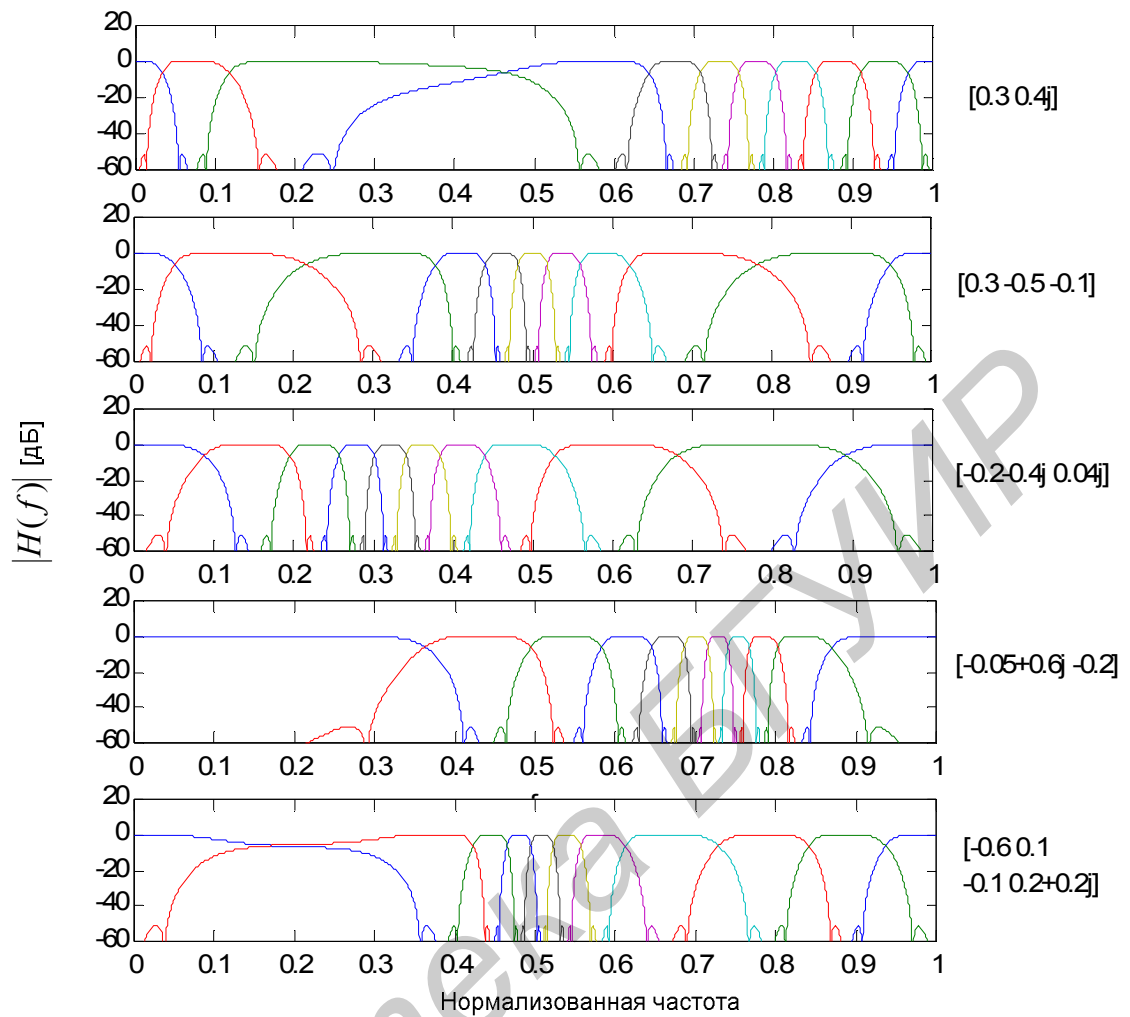


Рис. 1.20. Разложение субполос в банках с фазовыми фильтрами высших порядков

Другие области применения косинусных банков подобны ДПФ-модулированным. Часто применяемым косинусно-модулированным банком с неравнополосными каналами является декомпозиция на субполосы, согласно шкале Барков, которые согласованы с психоакустической моделью слуха человека.

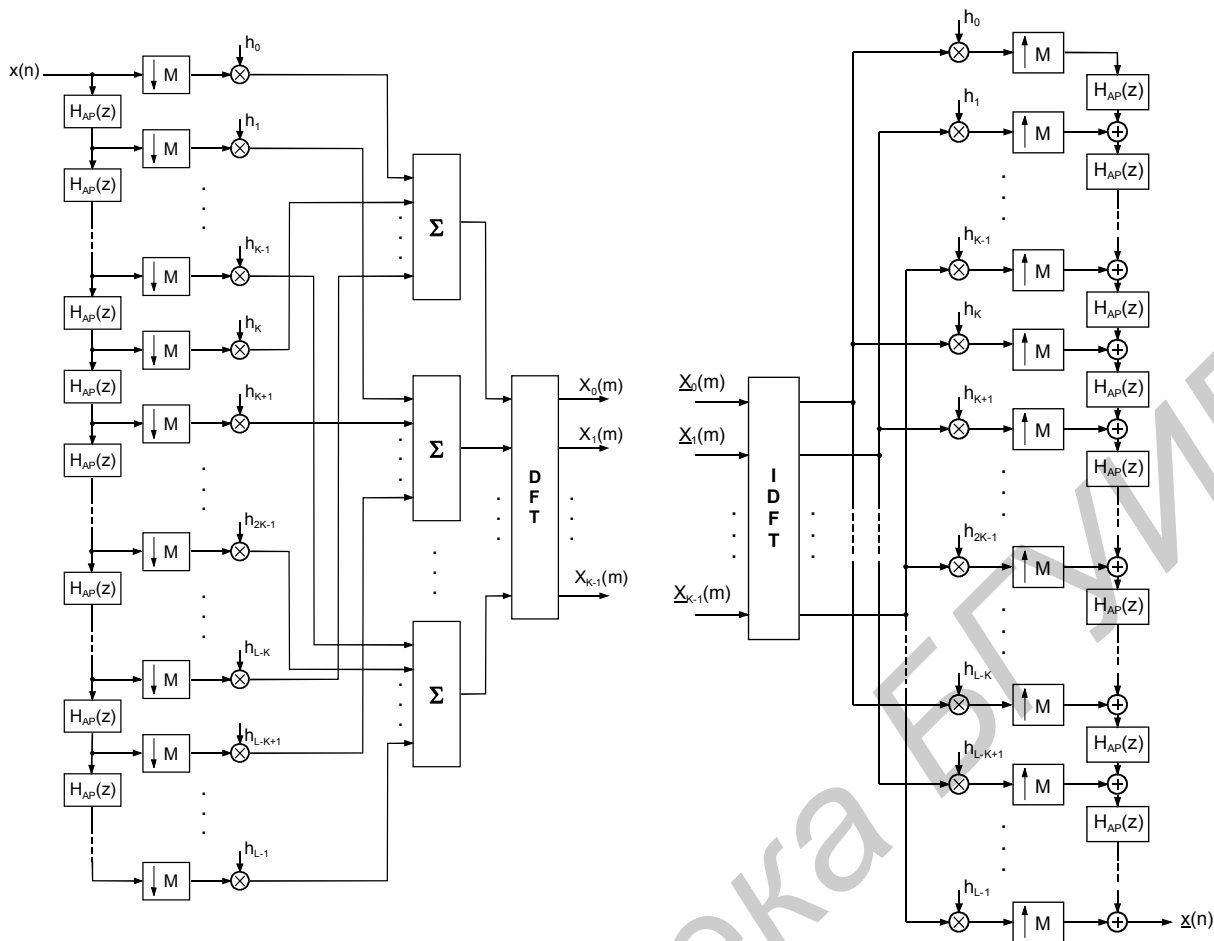


Рис. 1.21. ДПФ-модулированный неравнополосный банк фильтров анализа-синтеза

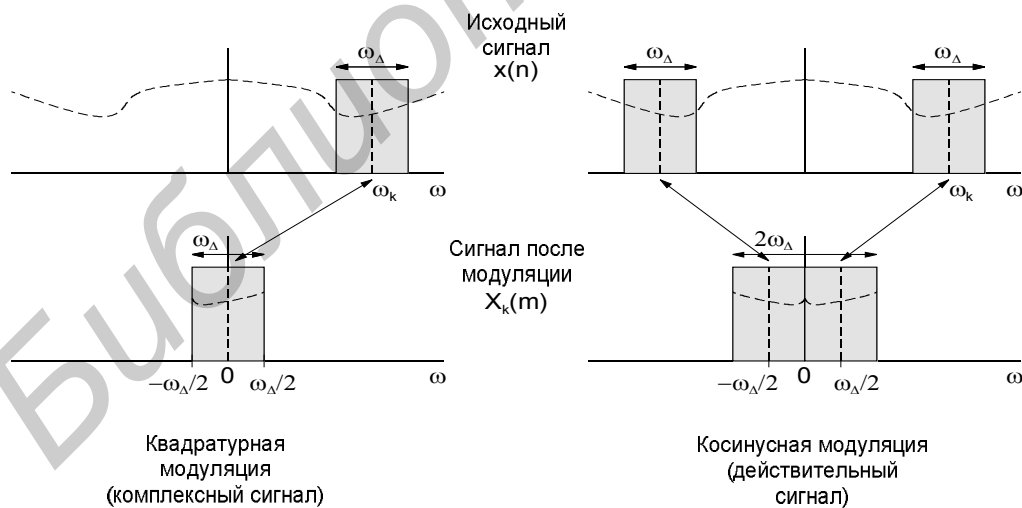


Рис. 1.22. Применение модуляции в банках фильтров

Реализации в среде MATLAB ДПФ-модулированных банков фильтров представлены ниже.

### 1.2.3. Реализация ДПФ-модулированных банков фильтров в среде MATLAB

Прямая реализация с полной модуляцией:

```
function [X, XLP, XM] = dft01a(x, K, M, h)
%DПФ-модулированный банк фильтров - анализатор
%Вход:
% x - анализируемый сигнал
% K - количество каналов
% M - коэффициент децимации
% h - коэффициенты фильтра-прототипа (ИХ) Low Pass FIR (fc =
% pi / K);
%Выход:
% X - матрица с сигналами каналов банка фильтра
% XM - сигналы каналов после модуляции
% XLP - сигналы каналов после фильтрации
%Формирование сигнала модуляции (гетеродин)
WnK = exp(-j * 2 * pi * (0 : (length(x) - 1)) / K);
%модуляция входного сигнала - получение канальных сигналов
X = zeros(K, length(x));
X(1, :) = x;
for i = 2 : K,
    X(i, :) = X(i - 1, :) .* WnK;
end
if (nargout > 2), XM = X; end
%НЧ-фильтрация каналов
for i = 1 : K,
    X(i, :) = filter(h, 1, X(i, :));
end
if (nargout > 2), XLP = X; end
%Децимация канальных сигналов
X = X(:, 1 : M : length(x));
```

```
function [x, xM, xLP, xZP] = dft01s(X, M, h)
%DПФ-модулированный банк фильтров - синтезатор
%Вход:
% X - матрица с сигналами каналов
% M - коэффициент интерполяции
% h - коэффициенты фильтра-прототипа (ИХ) Low Pass FIR (fc =
% pi / K)
%Выход:
% x - выходной сигнал
% xM - сигналы каналов после модуляции
% xLP - сигналы каналов после НЧ-фильтрации
% xZP - сигналы каналов после добавления нулевых отсчетов
%Определение количества каналов по размеру матрицы с сигналами каналов
K = size(X, 1);
%Добавление нулевых отсчетов
x(:, 1 : M : (M * size(X, 2))) = X;
if (nargout > 3), xZP = x; end
%Подготовка фильтра
h = h * M;
%Фильтрация каналов
for i = 1 : K,
    x(i, :) = filter(h, 1, x(i, :));
end
if (nargout > 2), xLP = x; end;
%Формирование коэффициентов модулятора (гетеродина)
WnK = exp(j * (0 : size(x, 2) - 1) * 2 * pi / K);
tmp = 1;
```

```

%Модуляция
for i = 2 : K,
    tmp = tmp .* WnK;
    x(i, :) = x(i, :) .* tmp;
end
if (nargout > 2),
    xM = x;
end;
%Суммирование выходов всех каналов и получение синтезированного сигнала
x = sum(x, 1);

```

### Максимально децимированный, полифазная реализация:

```

function [X, XF, XK, hp] = dft02a(x, K, h)
%Максимально децимированный ДПФ-модулированный банк фильтров - анализатор
%Вход:
% x - анализируемый сигнал
% K - количество каналов
% h - коэффициенты фильтра-прототипа (ИХ) Low Pass FIR (fc =
% pi / K)
%Выход:
% X - матрица с сигналами каналов банка фильтра после ДПФ
% XF - сигналы каналов после полифазной фильтрации
% XK - сигналы каналов после децимации
% hp - матрица полифазных фильтров
%Сортировка входных отсчетов сигнала на каналы - децимация
X = x;
if rem(length(X), K),
    X(K * ceil(length(X) / K)) = 0;
end
X = reshape(X, K, length(X) / K);
if (nargout > 2),
    XK = X;
end
%Разбиение фильтра-прототипа на полифазные фильтры
if mod(length(h), K),
    h(K * ceil(length(h) / K)) = 0;
end
h = flipud(reshape(h, K, length(h) / K));
if (nargout > 3),
    hp = h;
end
%Полифазная фильтрация каналов
for i = 1 : K,
    X(i, :) = filter(h(i, :), 1, X(i, :));
end
if (nargout > 2),
    XF = X;
end
%Модуляция посредством ДПФ
X = fft(X);

```

```

function [x, xF, xIDFT, hp] = dft02s(X, h)
%Максимально децимированный ДПФ-модулированный банк фильтров - синтезатор
%Вход:
% X - матрица с сигналами каналов
% h - коэффициенты фильтра-прототипа (ИХ) Low Pass FIR (fc =
% pi / K)
%Выход:
% x - синтезированный сигнал
% xF - сигналы в каналах после фильтрации
% xIDFT - сигналы в каналах после ОДПФ
% hp - матрица полифазных фильтров
%Определение количества каналов
K = size(X, 1);
%ОДПФ
x = ifft(X);
if (nargout > 2),
    xIDFT = x;
end
%Разбиение фильтра-прототипа на полифазные фильтры
if rem(length(h), K),

```



```

    h(K * ceil(length(h) / K)) = 0;
end
h = reshape(h, K, length(h) / K) * K * K;
if (nargout > 3),
    hp = h;
    %Полифазная фильтрация каналов
    for i = 1 : K,
        x(i, :) = filter(h(i, :), 1, x(i, :));
    end
    if (nargout > 1),
        xF = x;
        %Синтезированный сигнал
    x = x(:);
end
end

```

### Передискретизированный, полифазная реализация:

```

function [X, XF, XK] = dft03a(x, K, M, h)
%Передискретизированный ДПФ-модулированный банк фильтров - анализатор
%Вход:
% x - анализируемый сигнал
% M - коэффициент децимации фильтра
% I - коэффициент передискретизации каналов банка фильтра
% h - коэффициенты фильтра-прототипа Low Pass FIR (fc = pi/M)
%Выход:
% X - матрица с сигналами каналов
% XF - сигналы каналов после НЧ-фильтрации
% XK - сигналы каналов после децимации
%Определение количества каналов
I = K / M;
%Разделение отсчетов на каналы - децимация
X = x;
if mod(length(X), K),
    X(K * ceil(length(X) / K)) = 0;
end
X = reshape(X, K, length(X) / K);
if (nargout > 2),
    XK = X;
    %Разбиение фильтра-прототипа на полифазные фильтры
    if mod(length(h), K),
        h(M * ceil(length(h) / M)) = 0;
    end
    h = flipud(reshape(h, M, length(h) / M));
    if (nargout > 3),
        hp = h;
        %осуществление разреженного сбора фильтров
        h = repmat(h, I, 1);
        %Полифазная фильтрация с интерполяцией
        XF = zeros(K, size(X, 2) * I);
        for i = 1 : K,
            if 0,
                hi = reshape(h(i, :), I, length(h(i, :)) / I);
                z = floor((i - 1) / M);
                hi = flipud(hi);
                tmp = zeros(I, size(X, 2));
                for k = 1 : I;
                    tmp(k, :) = filter(hi(k, :), 1, [zeros(1, z) X(i, 1:size(X, 2)-z)]);
                end;
                XF(i, :) = tmp(:)';
            else
                k = 1 + floor((i - 1) / M);
                XF(i, k : I : length(XF)) = X(i, :);
                XF(i, :) = filter(h(i, :), 1, XF(i, :));
            end
        end
    end;
    X = XF;
end

```

```

%DПФ
X = fft(X);

function [x, xF, xIDFT] = dft03s(X, M, h)
%Передискретизированный ДПФ-модулированный банк фильтров - синтезатор
%Вход:
% X - матрица с сигналами каналов
% I - коэффициент передискретизации каналов банка фильтров
% h - коэффициенты фильтра-прототипа Low Pass FIR (fc = pi/M)
%Выход:
% x - синтезированный сигнал
% xF - сигналы каналов после фильтрации
% xIDFT - сигналы каналов после ОДПФ
%DПФ
x = ifft(X);
if (nargout > 2), xIDFT = x; end
%Определение коэффициента децимации фильтра
K = size(x, 1);
I = K / M;
%Разбиение фильтра-прототипа на полифазные фильтры
if mod(length(h), K) == 0;
    h(K * ceil(length(h) / K)) = 0;
end
h = reshape(h, M, length(h) / M * K * M);
if (nargout > 3), hp = h; end
%Осушение разреженного сбора фильтров
h = repmat(h, I, 1);
%Полифазная фильтрация каналов с последующей децимацией через I
xF = zeros(K, size(x, 2) / I);
for i = 1 : K,
    k = 1 + floor((i - 1) / M);
    %формирование сигнала согласно данного подмножества фильтров
    tmp = filter(h(i, :), 1, x(i, :));
    xF(i, :) = tmp(k : I : length(tmp));
end
%Синтезированный сигнал
x = xF(:);

```

### Обобщенный неравнополосный, полифазная реализация:

```

function [X] = dft04a(x, K, M, h, a, sm, sc)
%Обобщенный ДПФ-модулированный неравнополосный банк фильтров - анализатор
%Вход:
% x - анализируемый сигнал
% K - количество каналов
% M - коэффициент децимации канальных сигналов
% h - коэффициенты фильтра-прототипа Low Pass FIR (fc = pi/K)
% a - коэффициенты фазового фильтра
% sm - признак однократного перекрытия «1» или «0»
% sc - компенсация перекрытия полос «1» или «0»
%Выход:
% X - матрица с сигналами каналов банка фильтра
if ( nargin < 7 ), sc = 1; end
%Расчет коэффициентов всепропускающего фильтра
ap_a = 1;
ap_b = 1;
ap_K = length(a);
for i = 1 : ap_K,
    ap_a = conv(ap_a, [1 -a(i)]);
    ap_b = conv(ap_b, [-conj(a(i)) 1]);
    if sc,
        ap_abs = abs(a(i));
    end
end

```

```

        ap_ang = angle(a(i));
        ap_fi = -2*atan(ap_abs*sin(ap_ang) / (1-ap_abs*cos(ap_ang)));
        ap_b = ap_b * exp(j * ap_fi);
    end;
end;
%Подготовка фильтра-прототипа
lp_L = length(h);
lp_h = h * M * K;
%Фильтрация сигнала
if sm,
    X = zeros(K, ceil((length(x) + (lp_L - 1) * (ap_K - 1)) / M));
    tmp = x;
    for i = 1 : lp_L,
        qq = (lp_L - 1) * (ap_K - 1);
        pp = (lp_L - i) * (ap_K - 1);
        tmp2 = [zeros(1, pp) tmp zeros(1, qq - pp)];
        tmp2 = tmp2(1 : M : length(tmp2)) * lp_h(i);
        X(mod(i - 1, K) + 1, :) = X(mod(i - 1, K) + 1, :) + tmp2;
        tmp = filter(ap_b, ap_a, tmp);
    end
else
    X = zeros(K, ceil(length(x) / M));
    tmp = x;
    for i = 1 : lp_L,
        tmp2 = tmp(1 : M : length(tmp)) * lp_h(i);
        X(mod(i - 1, K) + 1, :) = X(mod(i - 1, K) + 1, :) + tmp2;
        tmp = filter(ap_b, ap_a, tmp);
    end
end
end
%Модуляция при помощи ОДПФ
X = ifft(X);

```

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

%Обобщенный ДПФ-модулированный неравнополосный банк фильтров - синтезатор

%Вход:

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

%Выход:

% x - синтезированный сигнал

if (nargin < 6), sc = 1; end

%Определение количества каналов

K = size(X, 1);

%Модуляция при помощи ДПФ

X = fft(X);

%Расчет коэффициентов всепропускающего фильтра

ap\_a = 1;

ap\_b = 1;

ap\_K = length(a);

for i = 1 : ap\_K,

ap\_a = conv(ap\_a, [1 -a(i)]);

ap\_b = conv(ap\_b, [-conj(a(i)) 1]);

if sc,

ap\_abs = abs(a(i));

ap\_ang = angle(a(i));

ap\_fi = -2 \* atan(ap\_abs \* sin(ap\_ang) / ...  
 (1 - ap\_abs \* cos(ap\_ang)));

ap\_b = ap\_b \* exp(j \* ap\_fi);

```

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

```

#### 1.2.4. Пример банка ДПФ-модулированных полифазных фильтров

Банк синтезирован на основе фильтра прототипа с конечной импульсной характеристикой. Число коэффициентов фильтра  $N=640$ . Число каналов в банке 64, т.е. коэффициент децимации (в анализирующем банке фильтров) и интерполяции (в синтезирующем банке фильтров) равен 64. На рис. 1.23 показана амплитудно-частотная характеристика фильтра прототипа для банка ДПФ-модулированных полифазных фильтров (анализа и синтеза).

При частоте дискретизации  $f_s=16\text{кГц}$  получается, что частота дискретизации в канале  $f_s/M=250\text{Гц}$ , полоса пропускания канала составляет 125Гц. На рис. 1.24 приведена амплитудно-частотная характеристика ДПФ-модулированного полифазного банка фильтров анализа-синтеза.

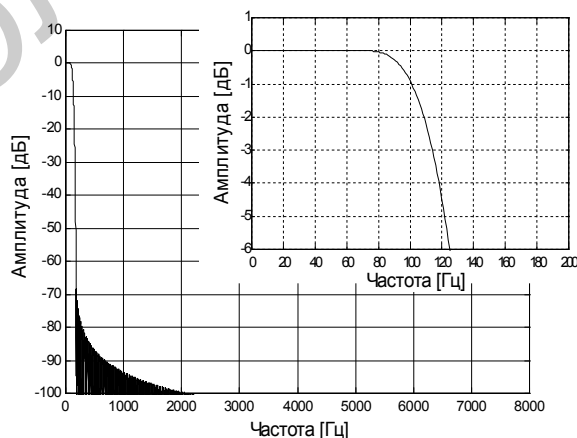


Рис. 1.23. Амплитудно-частотная характеристика фильтра прототипа для банка полифазных фильтров (анализа и синтеза)

Из рис. 1.24 видно, что амплитудных искажений в реконструированный сигнал (сигнал на выходе синтезирующего банка фильтров) не вносится. Перфективность реконструкции речевого сигнала иллюстрируется на рис. 1.25.

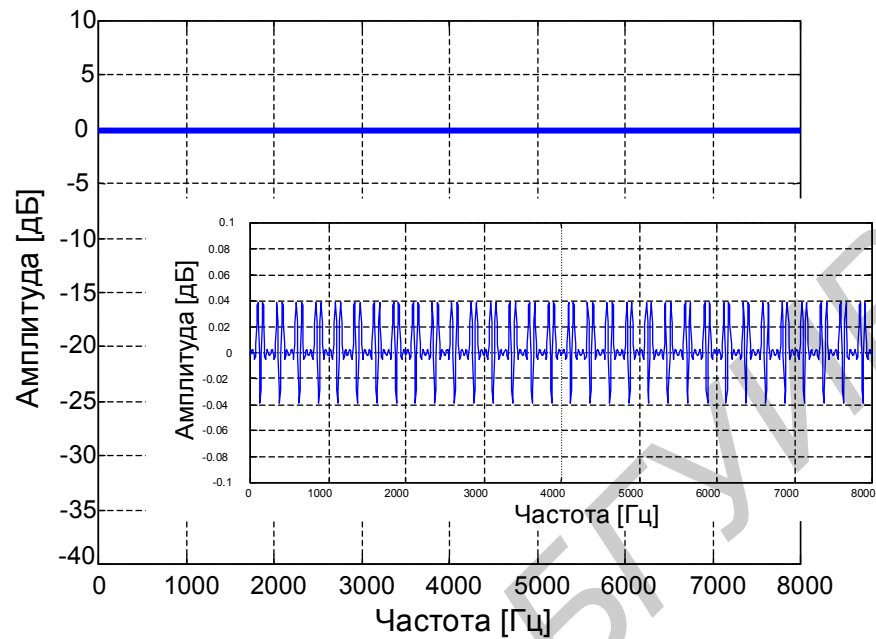


Рис. 1.24. Амплитудно-частотная характеристика полифазного банка фильтров анализа – синтеза

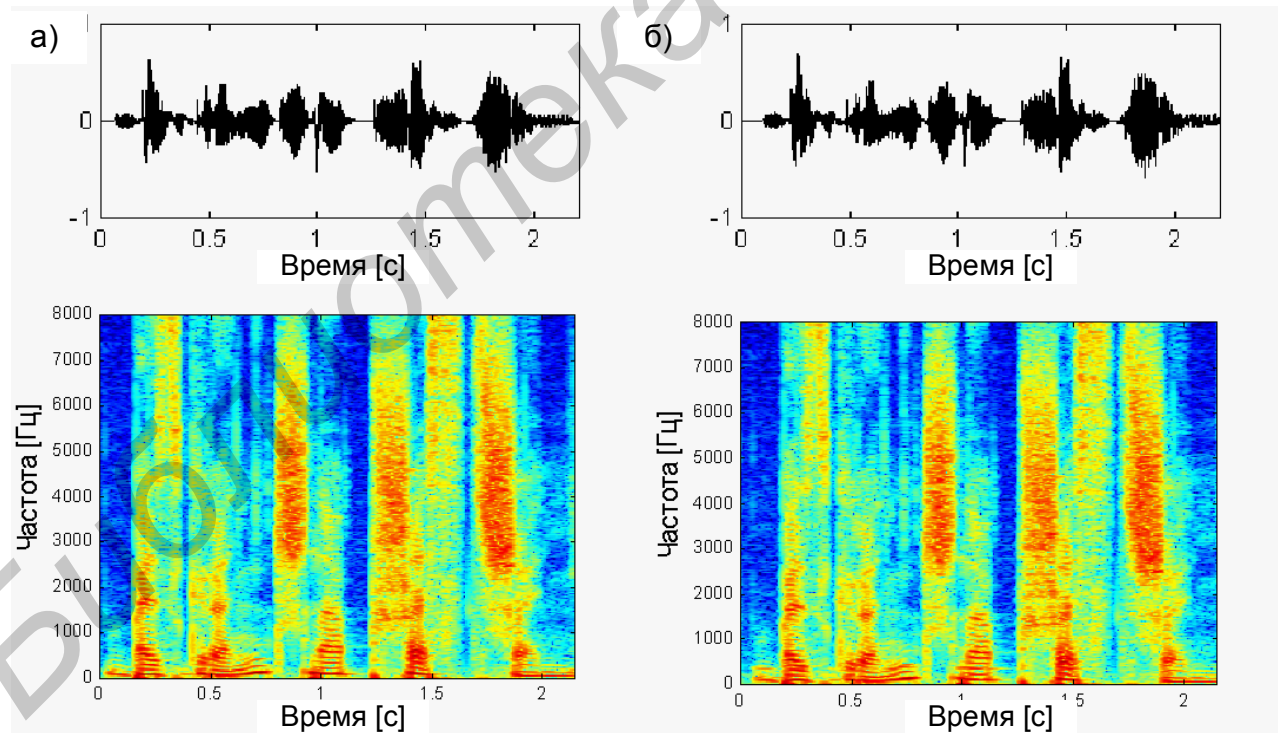


Рис. 1.25. Перфективная реконструкция входного речевого сигнала  $x(n/f_s)$  (а) и выходного  $\bar{x}((n - n_0)/f_s)$  (б), где  $n_0$  - алгоритмическая задержка алгоритма банков анализирующих и синтезирующих фильтров.

### 1.3. Косинусно-модулированные банки фильтров

#### 1.3.1. Проектирование косинусно-модулированного банка фильтров

В  $M$ -канальном косинусно-модулированном банке фильтров, фильтры анализа и синтеза в общем случае описываются следующими выражениями:

$$h_k(n) = 2h(n) \cos \left[ \frac{\pi}{M} \left( k + \frac{1}{2} \right) n + \varepsilon_k \right], \quad (1.23)$$

$$f_k(n) = 2h(n) \cos \left[ \frac{\pi}{M} \left( k + \frac{1}{2} \right) n + \gamma_k \right],$$

для  $k = 0, \dots, M-1$ , где  $h(n)$  и  $f(n)$  - НЧ КИХ-фильтры-прототипы с частотой среза  $\frac{\pi}{2M}$ . Выбор фаз  $\varepsilon_k$  и  $\gamma_k$  может быть осуществлен несколькими способами. Для обеспечения перфективной реконструкции фазы можно вычислить по выражениям, представленным ниже:

$$\varepsilon_k = -\frac{\pi}{M} \left( k + \frac{1}{2} \right) \frac{\alpha}{2} \quad \gamma_k = +\frac{\pi}{M} \left( k + \frac{1}{2} \right) \frac{\alpha}{2}. \quad (1.24)$$

В банке параунитарных фильтров принимают

$$\varepsilon_k = -\frac{\pi}{M} \left( k + \frac{1}{2} \right) \frac{L + M - 1}{2} + \frac{\pi}{2}, \quad (1.25)$$

$$\gamma_k = -\frac{\pi}{M} \left( k + \frac{1}{2} \right) \frac{L - M - 1}{2} - \frac{\pi}{2},$$

и, тем самым, используют фильтр-прототип  $h(n) = f(n)$  с длиной импульсной характеристики  $L = 2Mt$ , как на стороне анализа, так и на стороне синтеза.

Подставляя выражения (1.25) в (1.23), получим

$$h_k(n) = 2h(n) \cos \left[ \frac{\pi}{M} \left( k + \frac{1}{2} \right) \left( n - \frac{N-1}{2} \right) + (-1)^k \frac{\pi}{4} \right] \quad (1.26)$$

$$f_k(n) = 2h(n) \cos \left[ \frac{\pi}{M} \left( k + \frac{1}{2} \right) \left( n - \frac{N-1}{2} \right) - (-1)^k \frac{\pi}{4} \right]$$

Быстрая реализация косинусной модуляции может быть осуществлена путем использования БПФ при незначительной модификации передаточных функций фильтров-прототипов:

$$H_k(z) = a_k b_k U_k(z) + a_k^* b_k^* U_k^*(z), \quad (1.27)$$

$$F_k(z) = a_k^* b_k V_k(z) + a_k b_k^* V_k^*(z),$$

где

$$U_k(z) = H\left(zW^{k+\frac{1}{2}}\right), \quad V_k(z) = F\left(zW^{k+\frac{1}{2}}\right) \quad (1.28)$$

и

$$a_k = e^{j(-1)^k \frac{\pi}{4}}, \quad b_k = W^{-\frac{1}{2}\left(k+\frac{1}{2}\right)}. \quad (1.29)$$

Модуляция реализуется так, чтобы достичь наименьшего перекрытия между соседними субполосами. Перфективная реконструкция может быть достигнута соответствующим проектированием фильтра-прототипа. Ключом для этого является запись полифазной передаточной матрицы для банка анализа:

$$E(z) = \hat{C} \begin{bmatrix} g_0(-z^2) \\ z^{-1} g_1(-z^2) \end{bmatrix}, \quad (1.30)$$

где

$$g_0(z) = \text{diag}[G_0(z), G_1(z), \dots, G_{M-1}(z)], \quad (1.31)$$

$$g_1(z) = \text{diag}[G_M(z), G_{M+1}(z), \dots, G_{2M-1}(z)]$$

и

$$\hat{c}_{kl} = 2 \cos\left(\frac{\pi}{M}\left(k + \frac{1}{2}\right)\left(l - \frac{N-1}{2}\right) + (-1)^k \frac{\pi}{4}\right). \quad (1.32)$$

### 1.3.2. Косинусно-модулированные банки фильтров с квазиполным восстановлением

Существует ряд случаев, когда требования к совершенной реконструкции не являются критичными. Часто это связано с тем, что ошибка реконструкции банка фильтров может быть значительно ниже или соизмерима с ошибкой модификации субполосных сигналов, обрабатываемых по какому-либо алгоритму.

Большинство известных методик проектирования прототипа для косинусно-модулированных банков с квазиполным восстановлением основаны на равенстве

$$T(e^{j\omega}) = \sum_{k=0}^{M-1} F_k(e^{j\omega}) H_k(e^{j\omega}) = e^{j\omega(N-1)} \sum_{k=0}^{2M-1} |H(e^{j(\omega-k\pi/M)})|^2, \quad (1.33)$$

характеризующем ограничение на линейную фазо-частотную характеристику и уменьшение эффекта «алайзинга».

Ограничением является линейная фазочастотная характеристика фильтра-прототипа и частота среза  $\pi/M$  или

$$|H(e^{j\omega})| \approx 0, \quad |\omega| > \frac{\pi}{M}. \quad (1.34)$$

Таким образом, фильтр-прототип должен быть спроектирован так, чтобы амплитудно-частотная характеристика  $|T(e^{j\omega})|$  была в приближении плоской (с минимальными пульсациями) для всех частот  $\omega$  или  $|T(e^{j\omega})| \approx 1$ . Предложено много методов решения этой задачи с различными целевыми функциями и соответствующими процедурами оптимизации коэффициентов. Например, при поиске коэффициентов фильтра-прототипа может использоваться минимизация следующей целевой функции:

$$\phi = \max_{\omega} \left\{ |H(e^{j\omega})|^2 + |H(e^{j(\omega-\pi/M)})|^2 - 1 \right\}, \quad (1.35)$$

вычисленной для  $\omega \in (0, \pi/M)$ .

Полифазная структура неравнополосного косинусно-модулированного банка фильтров анализа и синтеза представлена на рис. 1.26 и рис. 1.27, соответственно.



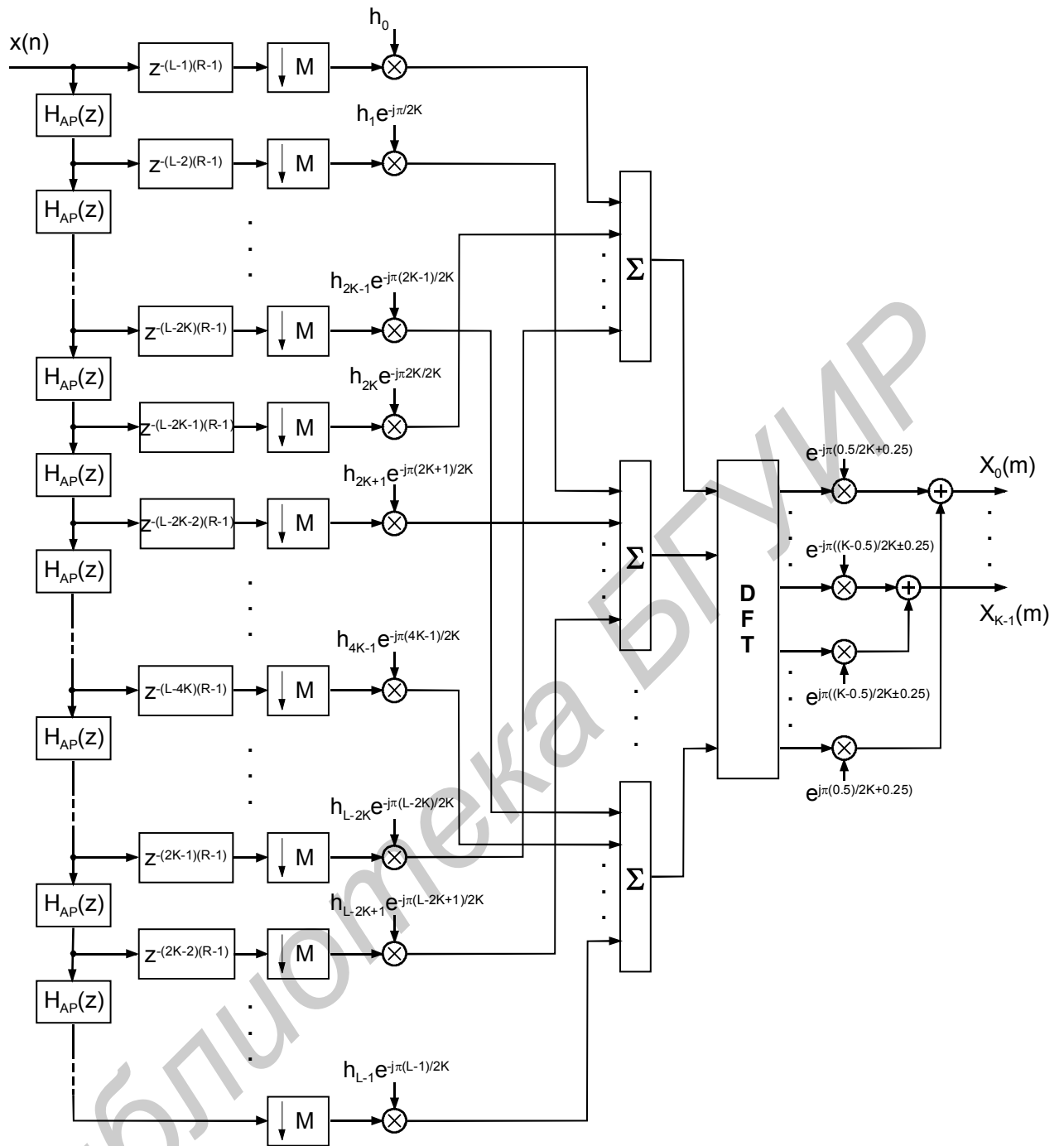


Рис. 1.26. Полифазная структура косинусно-модулированного неравнополосного банка фильтров анализа

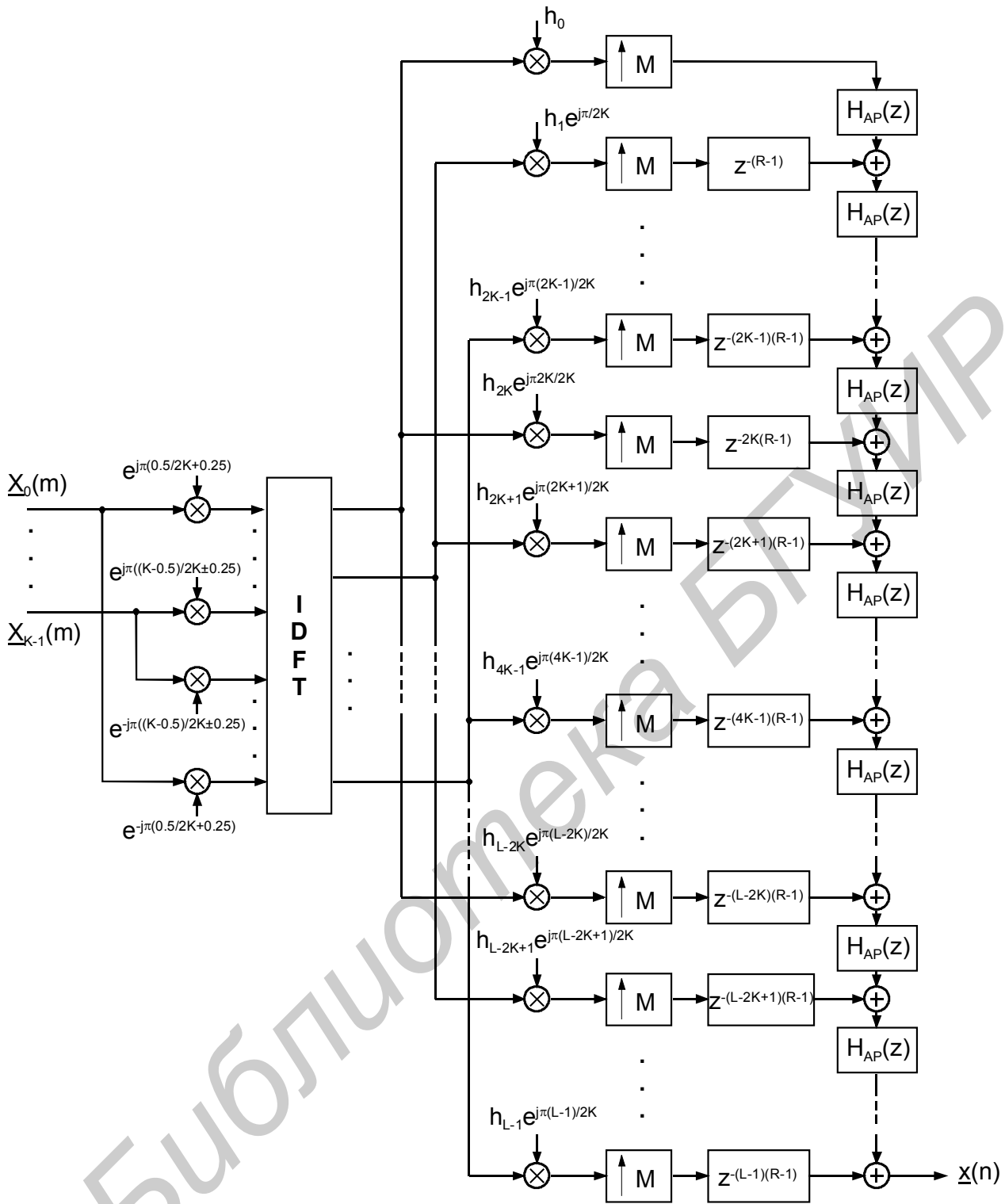


Рис. 1.27. Полифазная структура косинусно-модулированного неравнополосного банка фильтров синтеза

### 1.3.3. Реализация косинусно-модулированных банков фильтров в среде MATLAB

Прямая реализация:

```
function X = dct01a(x, K, M, lp_h)
%Косинусно-модулированный банк фильтров - анализатор
%Вход:
% x - анализируемый сигнал
% K - количество каналов
% M - коэффициент децимации
% lp_h - коэффициенты фильтра-прототипа LP FIR (fc = pi / (2*K))
%Выход:
% X - матрица с сигналами каналов банка фильтра
%Фильтр-прототип
lp_L = length(lp_h);
lp_h = lp_h * M;
%Модуляция фильтра - получение матрицы канальных фильтров
n = (0 : (lp_L - 1)) - ((lp_L - 1) / 2);
n = n * pi / K;
m = zeros(K, lp_L);
for k = [0 : (K - 1)],
    if mod(k, 2),
        m((k + 1), :) = (lp_h .* (2 * cos((1 * k + 0.5) * n + 1 * pi/4)));
    else
        m((k + 1), :) = lp_h .* (2 * cos((1 * k + 0.5) * n - 1 * pi/4));
    end
end;
%Фильтрация
X = zeros(K, length(x));
for k = 1 : K,
    X(k, :) = filter(m(k, :), 1, x);
end;
%Децимация канальных сигналов
X = X(:, 1 : M : size(X, 2));
```

```
function y = dct01s(X, M, lp_h)
%Косинусно-модулированный банк фильтров - синтезатор
%Вход:
% X - матрица с сигналами каналов банка фильтра
% M - коэффициент интерполяции
% lp_h - коэффициенты фильтра-прототипа LP FIR (fc = pi / (2*M))
%Выход:
% y - синтезированный сигнал
%Определение количества каналов
K = size(X, 1);
%Фильтр-прототип
lp_L = length(lp_h);
lp_h = lp_h * M;
%Модуляция фильтра - получение матрицы канальных фильтров
n = (0 : (lp_L - 1)) - ((lp_L - 1) / 2);
n = n * pi / K;
m = zeros(K, lp_L);
for k = [0 : (K - 1)],
    if mod(k, 2),
        m((k + 1), :) = (lp_h .* (2 * cos((1 * k + 0.5) * n - 1 * pi/4)));
    else
        m((k + 1), :) = lp_h .* (2 * cos((1 * k + 0.5) * n + 1 * pi/4));
    end
end;
n = m / M;
%Добавление нулевых отсчетов
```

```

Y = zeros(K, size(X, 2) * M);
Y(:, 1 : M : size(Y, 2)) = X;
%Фильтрация
y = 0;
for k = 1 : K,
    y = y + filter(n(k, :), 1, Y(k, :));
end;

```

### Равнополосный, максимально децимированный, параунитарный:

```

function [X] = dct02a(x, K, h)
%Равнополосный, максимально децимированный, параунитарный
%косинусно-модулированный банк фильтров - анализатор
%Вход:
% x - анализируемый сигнал
% K - количество каналов
% h - коэффициенты фильтра-прототипа LP FIR (fc = pi / (2*K))
%Выход:
% X - матрица сигналов каналов банка фильтра
%Подготовка фильтра-прототипа
if mod(length(h), 2 * K),
    h(2 * K * ceil(length(h) / (2 * K))) = 0;
end;
h = reshape(h, 2 * K, length(h) / (2 * K)) * K;
h(:, 2 : 2 : size(h, 2)) = -h(:, 2 : 2 : size(h, 2));
g = zeros(2 * K, 2 * size(h, 2));
g(1 : K, 1 : 2 : size(g, 2)) = h(1 : K, :);
g(K + 1 : 2 * K, 2 : 2 : size(g, 2)) = h(K + 1 : 2 * K, :);
if mod(length(x), K),
    x(K * ceil(length(x) / K)) = 0;
end;
%Сортировка входных отсчетов на каналы
X = repmat(reshape(x, K, length(x) / K), 2, 1);

%Полифазная фильтрация
for k = 1 : (2 * K),
    X(k, :) = filter(g(k, :), 1, X(k, :));
end;
%Подготовка модуляционной матрицы
for k = 1 : K,
    for l = 1 : (2 * K),
        c(k, l) = 2*cos((k-0.5) * (pi/K) * (l-0.5) + (-1 ^ (k-1)) * pi/4);
    end;
end;
%Модуляция
X = c * X;

function [y] = dct02s(X, h)
%Равнополосный, максимально децимированный, параунитарный
%косинусно-модулированный банк фильтров - синтезатор
%Вход:
% X - матрица с сигналами каналов банка фильтров
% h - коэффициенты фильтра-прототипа LP FIR (fc = pi / (2*K))
%Выход:
% y - синтезированный сигнал
%Определение количества каналов
K = size(X, 1);
%Подготовка фильтра-прототипа
if mod(length(h), 2 * K),
    h(2 * K * ceil(length(h) / (2 * K))) = 0;
end;
h = reshape(h, 2 * K, length(h) / (2 * K));

```

```

h(:, 1 : 2 : size(h, 2)) = -h(:, 1 : 2 : size(h, 2));
g = zeros(2 * K, 2 * size(h, 2));
g(1 : K, 1 : 2 : size(g, 2)) = h(1 : K, :);
g(K + 1 : 2 * K, 2 : 2 : size(g, 2)) = h(K + 1 : 2 * K, :);
%Формирование модуляционной матрицы
for k = 1 : K,
    for l = 1 : (2 * K),
        c(k, l) = 2*cos((k-0.5) * (pi/K) * (l-0.5) + (-1 ^ (k-1)) * pi/4);
    end;
end;
%Модуляция
Y = c' * X;
%Полифазная фильтрация
f = flipud(g);
for k = 1 : K,
    Y(k, :) = filter(f(k, :), 1, Y(k, :));
    Y(k + K, :) = filter(f(k + K, :), 1, Y(k + K, :));
end;
%Реконструкция сигнала
y = zeros(K, size(Y, 2));
for k = 1 : K,
    y(k, :) = Y(k, :) + Y(k + K, :);
end;
y = y(:);

```

### Обобщенный неравнополосный:

**function [X] = dct03a(x, K, M, h, a, sm, sc)**

%Обобщенный неравнополосный косинусно-модулированный банк фильтров -  
%анализатор

%Вход:

% x - анализируемый сигнал  
 % K - количество каналов  
 % M - коэффициент децимации канальных сигналов  
 % h - коэффициенты фильтра-прототипа Low Pass FIR ( $fc = \pi/2K$ )  
 % a - коэффициенты фазового фильтра  
 % sm - признак однократного перекрытия «1» или «0»  
 % sc - компенсация перекрытия полос «1» или «0»

%Выход:

% X - матрица с сигналами каналов банка фильтра

if (nargin < 7),

sc = 1;

if (nargin < 6), sm = 1; end

end

%Расчет коэффициентов всепропускающего фильтра

ap\_a = 1;

ap\_b = 1;

ap\_K = length(a);

for i = 1 : ap\_K,

ap\_a = conv(ap\_a, [1 -a(i)]);

ap\_b = conv(ap\_b, [-conj(a(i)) 1]);

if sc,

ap\_abs = abs(a(i));

ap\_ang = angle(a(i));

ap\_fi = -2\*atan(ap\_abs\*sin(ap\_ang) / (1-ap\_abs\*cos(ap\_ang)));

ap\_b = ap\_b \* exp(j \* ap\_fi);

end;

end;

%Подготовка фильтра-прототипа

lp\_L = length(h);

lp\_h = h \* M;

%Модуляция фильтра-прототипа

```

for k = 1 : lp_L,
    lp_h(k) = lp_h(k) * exp(-j * pi * (k - 1) / (2 * K));
end;
%Фильтрация
if sm,
    X = zeros(2 * K, ceil((length(x) + (lp_L - 1) * (ap_K - 1)) / M));
    tmp = x;
    for i = 1 : lp_L,
        qq = (lp_L - 1) * (ap_K - 1);
        pp = (lp_L - i) * (ap_K - 1);
        tmp2 = [zeros(1, pp) tmp zeros(1, qq - pp)];
        tmp2 = tmp2(1 : M : length(tmp2)) * lp_h(i);
        X(mod(i - 1, 2 * K) + 1, :) = X(mod(i - 1, 2 * K) + 1, :) + tmp2;
        tmp = filter(ap_b, ap_a, tmp);
    end;
else
    X = zeros(2 * K, length(x));
    tmp = x;
    for k = 1 : lp_L,
        X(mod(k-1, 2*K)+1, :) = X(mod(k-1, 2*K)+1, :) + tmp * lp_h(k);
        tmp = filter(ap_b, ap_a, tmp);
    end;
    X = X(:, 1 : M : size(X, 2));
end
%Модуляция
X = fft(X);
for k = 1 : 1 * K,
    tmp = exp(-j * pi * (((k - 0.5) / (2 * K)) - ((-1) ^ k) * 0.25));
    X(k, :) = X(k, :) * tmp;
    X(2 * K - k + 1, :) = X(2 * K - k + 1, :) * conj(tmp);
end;
X = X(1 : K, :) + flipud(X(K + 1 : 2 * K, :));

```

### **function [x] = dct03s(X, M, h, a, sm, sc)**

```

%Обобщенный неравнополосный косинусно-модулированный банк фильтров -
%синтезатор
%Вход:
% X - матрица с сигналами каналов
% M - коэффициент децимации каналов банка фильтров
% h - коэффициенты фильтра-прототипа Low Pass FIR (fc = pi/2K)
% a - коэффициенты фазового фильтра
% sm - признак однократного перекрытия «1» или «0»
% sc - компенсация перекрытия полос «1» или «0»
%Выход:
% x - синтезированный сигнал
if (nargin < 6),
    sc = 1;
    if (nargin < 5),
        sm = 1;
    end
end
%Расчет коэффициентов всепропускающего фильтра
ap_a = 1;
ap_b = 1;
ap_K = length(a);
for i = 1 : ap_K,
    ap_a = conv(ap_a, [1 -a(i)]);
    ap_b = conv(ap_b, [-conj(a(i)) 1]);
    if sc,
        ap_abs = abs(a(i));
        ap_ang = angle(a(i));
        ap_fi = -2*atan(ap_abs*sin(ap_ang) / (1-ap_abs*cos(ap_ang)));
        ap_b = ap_b * exp(j * ap_fi);
    end;
end;

```

```

end;
%Определение количества каналов
K = size(X, 1);
%Подготовка фильтра-прототипа
lp_L = length(h);
lp_h = 2 * K * h;
%Модуляция
X = [X; flipud(X)];
for k = 1 : 1 * K,
    tmp = exp(j * pi * ((k - 0.5) / (2 * K)) - ((-1) ^ k) * 0.25));
    X(k, :) = X(k, :) * tmp;
    X(2 * K - k + 1, :) = X(2 * K - k + 1, :) * conj(tmp);
end
X = ifft(X);
%Модуляция фильтра-прототипа
for k = 1 : lp_L,
    lp_h(k) = lp_h(k) * exp(j * pi * (k - 1) / (2 * K));
end
%Фильтрация
if sm,
    tmp1(1 : M : size(X, 2) * M) = X(1, :) * lp_h(1);
    x = [tmp1 zeros(1, (lp_L - 1) * (ap_K - 1))];
    for i = 2 : lp_L,
        x = filter(ap_b, ap_a, x);
        qq = (lp_L - 1) * (ap_K - 1);
        pp = (lp_L - i) * (ap_K - 1);
        tmp1(1:M:size(X,2)*M) = X(mod(i-1,2*K)+1,:) * lp_h(i);
        x = x + [zeros(1, qq - pp) tmp1 zeros(1, pp)];
    end
else
    x = zeros(1, size(X, 2) * M);
    x(1 : M : size(x, 2)) = filter(lp_h(1), 1, X(1, :));
    for i = 2 : lp_L,
        x = filter(ap_b, ap_a, x);
        x(1:M:length(x)) = x(1:M:length(x)) + X(mod(i-1,2*K)+1,:) * lp_h(i);
    end
end;
end;

```

## 2. Эффективность субполосного кодирования сигнала

### 2.1. Дисперсия ошибки реконструкции сигнала

Рассматривается субполосный кодер, представленный на рис. 1.1, состоящий из  $k = 0, 1, \dots, M - 1$  каналов. Каждый канал имеет анализирующий фильтр  $H_k(z)$ , синтезирующий фильтр  $F_k(z)$ , дециматор/интерполятор с коэффициентом передискретизации  $m_k$  и  $b_k$ -битный квантователь. Входной сигнал  $x[n]$  после пропускания через анализирующие фильтры становится  $x'_k[n]$ , а субполосные сигналы -  $x_k[n]$ . Субполосный кодер максимально децимирован, т.е. выполняется условие (2.1):

$$\sum_{k=0}^{M-1} m_k = 1. \quad (2.1)$$

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

Следовательно,  $S_{xx}(e^{jw})$  является спектральной плотностью мощности (СПМ) сигнала  $x[n]$ , а его дисперсия определяется как

$$\sigma_x^2 = \int_{-\pi}^{\pi} S_{xx}(e^{jw}) \frac{dw}{2\pi}. \quad (2.2)$$

После фильтрации анализирующим фильтром  $H_k(z)$ , дисперсия  $x'_k[n]$ , будет определяться выражением (2.3)

$$\sigma_x^2 = \int_{-\pi}^{\pi} S_{xx}(e^{jw}) |H_k(e^{jw})|^2 \frac{dw}{2\pi} \quad (2.3)$$

для  $k = 0, 1, \dots, M-1$ .

Так как сигнал стационарен в широком смысле, то дисперсия не изменяется после децимации, т.е. дисперсия  $x_k[n]$  субполосы  $k$  равна  $\sigma_k^2$ .

Квантователь моделируется как модель с аддитивным шумом, т.е. выходом квантователя является сигнал  $x_k[n] + q_k[n]$ , где  $x_k[n]$  входной сигнал и  $q_k[n]$  - шум квантования. Дисперсия шума квантования, как для равномерных квантователей, так и для квантователей, оптимизированных функцией плотности вероятности (ФПВ) определяется как

$$\sigma_{q_k}^2 = \varepsilon_k 2^{-2b_k} \sigma_k^2, \quad (2.4)$$

где  $\varepsilon_k$  - константа, которая зависит от функции плотности вероятности  $k$ -го субполосного сигнала.

## 2.2. Обобщенный банк фильтров

Рассмотрим рис. 1.1 (часть после квантования). Сигнал  $x_k[n] + q_k[n]$  поступает на каждый канал синтезирующего банка. Если поступают только субполосные сигналы  $x_k[n]$ , то банк обладает перфективной реконструкцией,



выход будет равен  $x[n]$ -входу (пренебрегая задержкой). Если поступают только шумы  $q_k[n]$ , соответствующие сигналы показаны на рис. 1.1 под сигнальными стрелками, сигналы  $q_k[n]$ , после интерполяции, становятся  $p_k[n]$ , которые после фильтрации становятся  $r_k[n]$  и выходной сигнал будет  $r[n]$ . Благодаря линейности синтезирующего банка, его выход, при поступлении сигнала и шумовой компоненты на вход, будет равен сумме этих двух выходов  $x[n] + r[n]$ . Поэтому, ошибка реконструкции (выход минус вход)  $r[n]$ , может быть полностью получена при фильтрации шума квантования синтезирующим банком фильтров.

Предположим, используются равномерные (многобитные) квантователи, тогда сигналы  $q_k[n]$  различных каналов некоррелированы между собой и являются «белым» шумом. Эта модель подходит для равномерных квантователей с количеством бит на временной отсчет более 3, и не подходит для равномерных квантователей с более низким количеством бит. СПМ сигналов  $q_k[n]$  является константой  $\sigma_{q_k}^2$ , так как это «белый» шум, и имеют нулевое среднее. Сигналы  $p_k[n]$ , полученные после интерполяции  $q_k[n]$ , определяются как

$$p_k[n] = \begin{cases} q_k[n/m_k], & \text{если } n \bmod m_k = 0 \\ 0 & \text{иначе} \end{cases}, \quad (2.5)$$

и больше не являются стационарными в широком смысле, но имеют циклическую стационарность с периодом  $m_k$ . Другими словами, СПМ сигнала  $p_k[n]$ , если  $n \bmod m_k = 0$ , равна  $\sigma_{q_k}^2$  и СПМ для оставшихся отсчетов равна нулю. Шумовые составляющие  $p_k[n]$  пропускаются через синтезирующий фильтр  $F_k(z)$ , на выходе которого получается сигнал  $r_k[n]$ , стационарный в широком смысле с периодом  $m_k$ . Усредненная на периоде  $m_k$  дисперсия сигнала  $r_k[n]$  определяется как

$$\frac{1}{m_k} \left[ \int_{-\pi}^{\pi} \sigma_{q_k}^2 |F_k(e^{jw})|^2 \frac{dw}{2\pi} + 0 + \dots + 0 \right], \quad (2.6)$$

где  $m_k - 1$  – нули, полученные из отсчетов сигнала с шумовой составляющей, равной нулю.

Ошибка реконструкции сигнала  $r[n]$  равна сумме всех  $r_k[n]$ . Так как  $r_k[n]$  во всех каналах некоррелированы, то дисперсия их суммы равна сумме их дисперсий:

$$\sigma_r^2 = \sum_{k=0}^{M-1} \frac{\sigma_{q_k}^2}{m_k} \int_{-\pi}^{\pi} |F_k(e^{jw})|^2 \frac{dw}{2\pi}, \quad (2.7)$$

где  $\sigma_r^2$  - дисперсия ошибки реконструкции.

Пусть  $n_k = \int_{-\pi}^{\pi} |F_k(e^{jw})|^2 \frac{dw}{2\pi}$  означает нормы синтезирующих фильтров (заметим, что для КИХ-фильтра с коэффициентами импульсной характеристики  $[h_0, h_1, \dots, h_L]$  норма равна  $h_0^2 + h_1^2 + \dots + h_L^2$ ). Затем, используя уравнение (2.4), выражение выше преобразуется к выражению (2.8):

$$\sigma_r^2 = \sum_{k=0}^{M-1} \frac{\varepsilon_k 2^{-2b_k} \sigma_k^2 n_k}{m_k}. \quad (2.8)$$

### 2.3. Ортогональный банк фильтров

#### 2.3.1. Дисперсия ошибки

Стороны анализа и синтеза на рис. 1.1 могут быть выражены в полифазной форме, как показано на рис. 2.1. Для параунитарных систем свойство отсутствия потерь (или энергетический баланс) говорит о том, что энергия выхода равна энергии входа. Рассматривается только шум квантования, поступающий на синтезирующий банк. Пусть  $\sigma_{q_i}^{\prime 2}$  будет дисперсией шума квантования  $i$ -го канала для  $i = 0, 1, \dots, N-1$ . Следовательно, дисперсия (энергия) на входе, усредненная по всем каналам будет равна  $\frac{1}{N} \sum_{i=0}^{N-1} \sigma_{q_i}^{\prime 2}$ , что равно усредненной дисперсии выхода после полифазного синтеза.

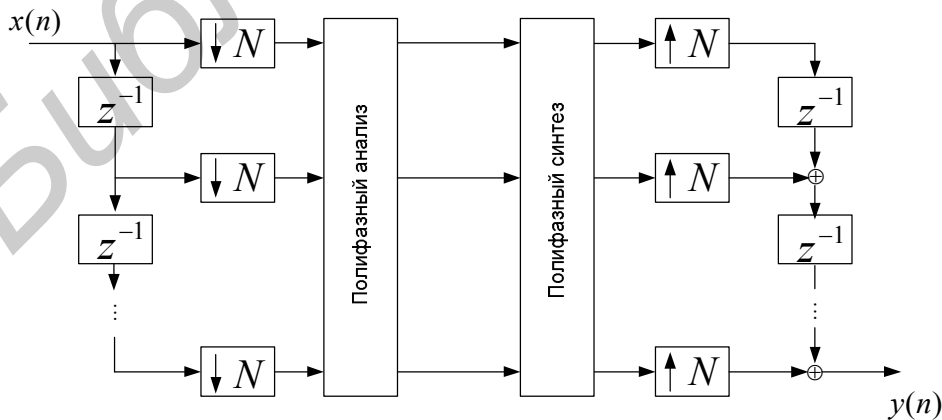


Рис. 2.1. Субполосный кодер в полифазной форме

Цепочка интерполяции, задержки и суммирования просто преобразует выходную векторную последовательность в скалярную последовательность. Поэтому, это эквивалентно усредненной дисперсии выхода или дисперсии ошибки реконструкции:

$$\sigma_r^2 = \frac{1}{N} \sum_{i=0}^{N-1} \sigma_{q_i}^{\prime 2}. \quad (2.9)$$

Заметим, что  $N$  каналов на рис. 2.1 получены из  $M$  каналов на рис. 1.1. Следовательно, первые  $\frac{N}{m_0}$  каналов рис. 2.1 имеют тот же квантователь, что и канал 0 на рис. 1.1, поэтому  $\sigma_{q_i}^{\prime 2}$  равна  $\sigma_{q_0}^2$  для  $i = 0, \dots, \frac{N}{m_0} - 1$ . Подобно могут быть приравнены другие дисперсии и выражены следующим образом:

$$\sigma_r^2 = \frac{1}{N} \sum_{k=0}^{M-1} \frac{N}{m_k} \sigma_{q_k}^2 = \sum_{k=0}^{M-1} \frac{\sigma_{q_k}^2}{m_k}. \quad (2.10)$$

Используя уравнение (2.4), выражение (2.10) преобразуется к виду (2.11)

$$\sigma_r^2 = \sum_{k=0}^{M-1} \frac{\varepsilon_k 2^{-2b_k} \sigma_k^2}{m_k}. \quad (2.11)$$

Сравнивая уравнения (2.8) и (2.11) для ортогонального случая можем заметить, что в выражении (2.11) отсутствует норма синтезирующего фильтра  $n_k$ . Это является следствием свойства отсутствия потерь, подразумевающего фильтры с единичной энергией или  $n_k = 1$  для всех  $k$ . Заметим также, что допущение об ортогональном банке фильтров заменяет допущение о равномерных (многобитных) квантователях и не является необходимым для получения данного результата. Поэтому, в смысле квантователя, результат более общий для ортогонального случая.

## **2.4. Оптимальное распределение бит в каналах банка фильтров**

### **2.4.1. Метод множителей Лагранжа**

Обращаясь к рис. 1.1,  $b_k$  является количеством бит на отсчет для  $k$ -го квантователя. Однако, благодаря коэффициенту децимации  $m_k$ , один отсчет квантуется этим квантователем для каждого из  $m_k$  входных отсчетов. Поэтому

скорость передачи квантователя равна  $\frac{b_k}{m_k}$  бит на входной отсчет.

Следовательно, усредненная скорость передачи субполосной системы можно определить как

$$b = \sum_{k=0}^{M-1} \frac{b_k}{m_k} \text{ бит/отсчет.} \quad (2.12)$$

Проблема оптимального распределения бит в каналах состоит в нахождении  $b_0, b_1, \dots, b_{M-1}$ , которые минимизируют дисперсию ошибки реконструкции  $\sigma_r^2$  в уравнении (2.8), удовлетворяя ограничению (2.12). Данная минимизация с ограничением (2.12) может быть решена, используя метод множителей Лагранжа. Для этого определяется целевая функция

$$C = \sum_{k=0}^{M-1} \frac{1}{m_k} \varepsilon_k 2^{-2b_k} \sigma_k^2 n_k + \lambda \left( \sum_{k=0}^{M-1} \frac{b_k}{m_k} - b \right), \quad (2.13)$$

где  $\lambda$  - множитель Лагранжа.

Дифференцируя  $C$  по  $b_k$  и приравнявая к нулю ( $\frac{\partial C}{\partial b_k} = 0$ )

$$\frac{1}{m_k} \varepsilon_k 2^{-2b_k} \sigma_k^2 n_k (-2 \ln 2) + \lambda \frac{1}{m_k} = 0, \quad (2.14)$$

получается равенство (2.15)

$$\frac{\partial}{\partial b_k} 2^{-2b_k} = \frac{\partial}{\partial b_k} e^{-2 \ln 2 b_k} = -2 \ln 2 e^{-2 \ln 2 b_k}. \quad (2.15)$$

Следовательно,

$$2 \ln 2 \varepsilon_k 2^{-2b_k} \sigma_k^2 n_k = \lambda \quad (2.16)$$

или

$$2^{-2b_k} = \frac{\lambda}{2 \ln 2 \varepsilon_k \sigma_k^2 n_k}. \quad (2.17)$$

Возьмем логарифм по основанию 2 от обеих частей:

$$-2b_k = \log_2 \frac{\lambda}{2 \ln 2 \varepsilon_k \sigma_k^2 n_k} \quad (2.18)$$

или

$$b_k = \frac{1}{2} \log_2 \frac{2 \ln 2 \varepsilon_k \sigma_k^2 n_k}{\lambda} = \frac{1}{2} \log_2 \frac{2 \ln 2}{\lambda} + \frac{1}{2} \log_2 (\varepsilon_k \sigma_k^2 n_k). \quad (2.19)$$

Выражение выше действительно для всех значений  $k$  от 0 до  $M-1$ . Подставляя уравнение (2.18) в ограничение (2.12), получается, что

$$b = \sum_{k=0}^{M-1} \frac{1}{m_k} \left[ \frac{1}{2} \log_2 \frac{2 \ln 2}{\lambda} + \frac{1}{2} \log_2 (\varepsilon_k \sigma_k^2 n_k) \right] \quad (2.20)$$

$$= \frac{1}{2} \log_2 \frac{2 \ln 2}{\lambda} \sum_{k=0}^{M-1} \frac{1}{m_k} + \frac{1}{2} \sum_{k=0}^{M-1} \log_2 (\varepsilon_k \sigma_k^2 n_k)^{1/m_k}. \quad (2.21)$$

Первая сумма равна 1 на основании уравнения (2.1). Вторая сумма логарифмов может быть записана как логарифм произведения. Используя эти упрощения, средняя скорость передачи в субполосной системе равна

$$b = \frac{1}{2} \log_2 \frac{2 \ln 2}{\lambda} + \frac{1}{2} \log_2 \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{1/m_i} \quad (2.22)$$

или

$$\frac{1}{2} \log_2 \frac{2 \ln 2}{\lambda} = b - \frac{1}{2} \log_2 \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{1/m_i}. \quad (2.23)$$

Подставляя результат, полученный выше, в уравнение (2.18), получается следующее распределение бит по каналам:

$$b_k = b - \frac{1}{2} \log_2 \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{1/m_i} + \frac{1}{2} \log_2 (\varepsilon_k \sigma_k^2 n_k) \quad (2.24)$$

$$= b + \frac{1}{2} \log_2 \frac{\varepsilon_k \sigma_k^2 n_k}{\prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{1/m_i}}. \quad (2.25)$$

для всех значений  $k$  от 0 до  $M - 1$ .

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

#### 2.4.2. Оптимальное распределение бит

Задача оптимального распределения бит состоит в нахождении  $b_0, b_1, \dots, b_{M-1}$  удовлетворяющих условию (ограничению)  $b = \sum_k b_k / m_k$ , которые минимизируют целевую функцию  $\sum_k \varepsilon_k 2^{-2b_k} \sigma_k^2 n_k / m_k$ .

Ограничение может быть записано как

$$2^{-2b} = \prod_k (2^{-2b_k})^{1/m_k}. \quad (2.26)$$

Обе части (2.26) умножаются на  $\prod_k (\varepsilon_k \sigma_k^2 n_k)^{1/m_k}$ , чтобы получить следующее выражение

$$2^{-2b} \prod_k (\varepsilon_k \sigma_k^2 n_k)^{1/m_k} = \prod_k (2^{-2b_k} \varepsilon_k \sigma_k^2 n_k)^{1/m_k}, \quad (2.27)$$

левая часть которого – константа, так как в ней отсутствует  $b_k$ . Правая часть выражается через  $X_k = \varepsilon_k 2^{-2b_k} \sigma_k^2 n_k$  (где  $X_k$  - переменные, так как являются функциями от  $b_k$ ) следующим образом:

$$const = \prod_k X_k^{1/m_k}. \quad (2.28)$$

Целевая функция также может быть упрощена, используя  $X_k$ . Итак, появляется задача найти  $X_0, \dots, X_{M-1}$ , удовлетворяющие (2.28), которые

минимизируют  $\sum_k X_k / m_k$ . Ограничение может быть далее выражено как

$const = \prod_{k=0}^{M-1} (X_k^{N/m_k})^{1/N}$ . Определим  $N$  переменных  $Y_i$ , таким образом, что первая

$\frac{N}{m_0} Y_i$  равна  $X_0$  и так далее. Теперь ограничение можем переписать как

$const = \prod_{i=0}^{N-1} (Y_i)^{1/N}$ . Возведя обе части в степень  $N$ , получается  $const = \prod_i Y_i$ .

Подобно определится целевая функция  $\sum_{k=0}^{M-1} \frac{1}{N} \cdot \frac{N}{m_k} X_k = \sum_{i=0}^{N-1} \frac{1}{N} Y_i$ , где постоянный

коэффициент  $\frac{1}{N}$  может быть опущен. Таким образом, задача заключается в

поиске  $Y_0, \dots, Y_{N-1}$ , удовлетворяющих условию  $\prod_i Y_i$ , которое минимизирует

$$\sum_i Y_i.$$

Так как  $X_k = \varepsilon_k 2^{-2b_k} \sigma_k^2 n_k \geq 0$ , то переменные  $Y_i$  неотрицательные. Решение, минимизирующее сумму нескольких неотрицательных переменных при постоянной величине их произведения, известно и выполняется, когда все переменные равны. Поэтому решением являются  $Y_0 = Y_1 = \dots = Y_{N-1}$  или  $X_0 = X_1 = \dots = X_{M-1}$  для некоторого значения  $X$ .

Физическая интерпретация этого решения – следующая: все  $X_k = \varepsilon_k 2^{-2b_k} \sigma_k^2 n_k$  равны и, следовательно, согласно уравнению (2.4), все  $\sigma_k^2 n_k$  равны или энергия шума квантования, наблюдаемого после синтезирующего банка, равна в каждом канале.

Значение  $X$  может быть найдено из ограничения. Так как  $X_k = X$ , то получается  $2^{-2b_k} = X / \varepsilon_k \sigma_k^2 n_k$ . После подстановки данного выражения в уравнение (2.26) следует, что

$$2^{-2b_k} = \prod_k \left( \frac{X}{\varepsilon_k \sigma_k^2 n_k} \right)^{\frac{1}{m_k}} = \frac{X^{\frac{1}{m_0} + \dots + \frac{1}{m_{M-1}}}}{\prod_k (\varepsilon_k \sigma_k^2 n_k)^{\frac{1}{m_k}}} \quad (2.29)$$

или, используя уравнение (2.1),

$$X = 2^{-2b} \prod_k (\varepsilon_k \sigma_k^2 n_k)^{\frac{1}{m_k}}. \quad (2.30)$$

## 2.5. Оценка эффективности субполосного кодирования

На основании всего вышеизложенного, находится дисперсия минимальной ошибки реконструкции, используя оптимальное распределение бит (уравнение (2.24)):

$$2^{-2b_k} = 2^{-2b} \frac{\prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{1/m_i}}{\varepsilon_k \sigma_k^2 n_k}, \quad (2.31)$$

для всех  $k$  от 0 до  $M-1$ .

Подставляя это выражение в уравнение (2.8) и принимая во внимание (2.1), получается, что

$$\sigma_r^2 = \sum_{k=0}^{M-1} \frac{1}{m_k} 2^{-2b} \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{1/m_i} \quad (2.32)$$

$$= 2^{-2b} \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{1/m_i} \sum_{k=0}^{M-1} \frac{1}{m_k} \quad (2.33)$$

$$= 2^{-2b} \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{1/m_i}, \quad (2.34)$$

Эффективность субполосного кодирования определяется как отношение дисперсии ошибки реконструкции полнополосного кодера (или импульсно-кодовой модуляции – ИКМ) к субполосному кодеру с аналогичной скоростью потока данных. Полнополосный кодер просто квантует вход  $x[n]$ , используя  $b$ -битный квантователь. Поэтому, из уравнения (2.4), его дисперсия ошибки квантования (которая также является дисперсией ошибки реконструкции) определяется как

$$\sigma_q^2 = \varepsilon 2^{-2b} \sigma_x^2, \quad (2.35)$$

где  $\varepsilon$  как и ранее. Следовательно, эффективность субполосного кодирования можно определить как

$$G = \frac{\varepsilon 2^{-2b} \sigma_x^2}{2^{-2b} \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{1/m_i}} = \frac{\varepsilon \sigma_x^2}{\prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i)^{1/m_i}}. \quad (2.36)$$



### 2.5.1. Сигнал с нормальным законом распределения

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

$$G = \frac{\varepsilon \sigma_x^2}{\left( \varepsilon^{\frac{1}{m_0}} + \dots + \varepsilon^{\frac{1}{m_{M-1}}} \right) \prod_{i=0}^{M-1} (\sigma_i^2 n_i)^{1/m_i}} = \frac{\sigma_x^2}{\prod_{i=0}^{M-1} (\sigma_i^2 n_i)^{1/m_i}}. \quad (2.37)$$

### 2.5.2. Банк фильтров с равными коэффициентами децимации

Пусть коэффициенты децимации будут равны  $m_0 = m_1 = \dots = m_{M-1}$ . Из уравнения (2.1) следует, что каждый коэффициент равен  $\frac{1}{M}$  и эффективность субполосного кодирования составит

$$G = \frac{\varepsilon \sigma_x^2}{\left[ \prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2 n_i) \right]^{\frac{1}{M}}}. \quad (2.38)$$

### 2.5.3. Ортогональный банк фильтров

В случае ортогонального банка фильтров  $n_k = 1$ , как было обозначено ранее, улучшение при субполосном кодировании составит

$$G = \frac{\varepsilon \sigma_x^2}{\prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2)^{1/m_i}}. \quad (2.39)$$

Далее, свойство отсутствия потерь, обозначенное ранее, может быть применено к матрице полифазного анализа на рис. 2.1 методом, подобным описанному в п.2.3.1. Так как входом полифазного анализа является  $x[n]$  (в векторной или параллельной форме), его усредненная дисперсия входа будет

равна  $\sigma_x^2$ . Пусть  $\sigma_i'^2$  обозначает дисперсию выхода  $i = 0, \dots, N-1$  субполос. Тогда из свойства отсутствия потерь следует:

$$\sigma_x^2 = \frac{1}{N} \sum_{i=0}^{N-1} \sigma_i'^2. \quad (2.40)$$

Так как  $N$  каналов на рис. 2.1 получены из  $M$  каналов на рис. 1.1, первые  $\frac{N}{m_0}$  каналов имеют такую же субполосную дисперсию как и канал 0 на рис. 1.1, поэтому  $\sigma_i'^2$  равна  $\sigma_0^2$  для  $i = 0, \dots, \frac{N}{m_0} - 1$ . Подобно могут быть приравнены и другие субполосные дисперсии:

$$\sigma_x^2 = \frac{1}{N} \sum_{k=0}^{M-1} \frac{N}{m_k} \sigma_k^2 = \sum_{k=0}^{M-1} \frac{\sigma_k^2}{m_k}. \quad (2.41)$$

Используя это равенство в уравнении (2.38), эффективность субполосного кодирования может быть выражена следующим образом:

$$G = \frac{\varepsilon \sum_{i=0}^{M-1} \sigma_i^2 / m_i}{\prod_{i=0}^{M-1} (\varepsilon_i \sigma_i^2)^{1/m_i}}. \quad (2.42)$$

#### 2.5.4. Входной сигнал с нормальным законом распределения и ортогональный банк фильтров с равными коэффициентами децимации

Наиболее упрощенное выражение для эффективности субполосного кодирования получается при одновременном рассмотрении следующих допущений: входной сигнал имеет нормальный закон распределения, коэффициенты децимации равны, ортогональный банк фильтров. При этом оно становится отношением арифметического среднего к геометрическому среднему неотрицательных величин  $\sigma_i^2$ :

$$G = \frac{\frac{1}{M} \sum_{i=0}^{M-1} \sigma_i^2}{\left( \prod_{i=0}^{M-1} \sigma_i^2 \right)^{\frac{1}{M}}}. \quad (2.43)$$

Так как арифметическое среднее больше чем или равно геометрическому среднему, следовательно  $G \geq 1$ , т.е. субполосный кодер обеспечивает равное или лучшее качество по сравнению с полнополосным кодером при любом входном сигнале. Выражение (2.43) также известно как *мера пологости спектра* (от англ. – «Spectrum Flatness Measure» – SFM). Часто, в задачах кодирования речи используется взвешенная на порог маскирования величина SFM (*перцептуально взвешенная SFM* – от англ. «Perceptually Weighted SFM» – PSFM).

### 3. ДПФ с неравномерным частотным разрешением (WDFT)

#### 3.1.1. Основные принципы WDFT

ДПФ с неравномерным разложением коэффициентов Фурье по частотной оси (от англ. – «Warped Discrete-Fourier Transform» - WDFT) позволяет получить неравномерное  $z$ -преобразование конечной последовательности, используя фазовый фильтр  $A(z)$ . WDFT последовательности  $x[n]$  из  $N$  точек определяется по следующей формуле:

$$\hat{X}(z_k) = X(\hat{z}_k) = \sum_{n=0}^{N-1} x[n] \hat{z}_k^{-n}, \quad k = 0, \dots, N-1, \quad (3.1)$$

где  $\hat{z}_k$ -изображения равноотстоящих точек на единичной окружности в  $z$ -плоскости, получаемые из преобразования

$$z_k^{-1} = e^{-j \frac{2\pi k}{N}} \rightarrow \hat{z}_k^{-1} = A(z_k) \quad k = 0, \dots, N-1 \quad (3.2)$$

с фазовым звеном  $A(z)$  произвольного порядка.

Простейший вариант WDFT основан на фазовом звене первого порядка с действительным коэффициентом (1.18) Условием стабильности является  $|a| < 1$ . В зависимости от знака  $a$ , растягивается низкочастотный ( $a > 0$ ) или высокочастотный ( $a < 0$ ) диапазон, таким образом, что оставшаяся часть единичной окружности становится сжатой. Формально это может быть выражено:

$$\hat{\omega} = \omega + 2 \arctan \left( \frac{a \sin \omega}{1 - a \cos \omega} \right) \quad \text{для} \quad \begin{cases} z = e^{j\omega} \\ \hat{z} = e^{j\hat{\omega}} \end{cases}, \quad (3.3)$$

то есть осуществляется билинейное преобразование  $z$ -плоскости в новую искривленную  $\hat{z}$ -плоскость (рис. 3.1).

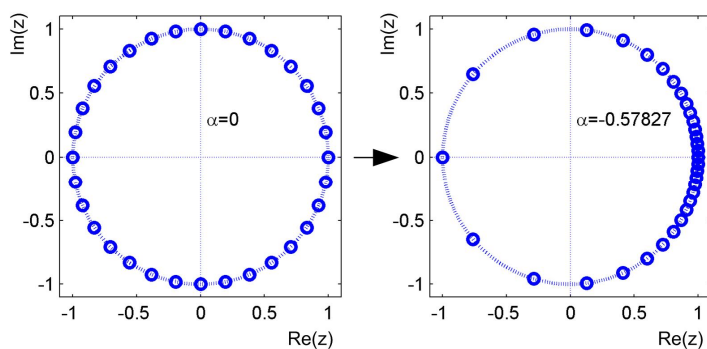


Рис. 3.1. Расположение частотных отсчетов ДПФ (слева) и WDFТ (справа)

Таким образом, WDFТ является обобщением ДПФ с частотными компонентами, расположенными неравномерно, но регулярно на единичной окружности. В матричной записи (с  $\hat{X}[k]$ , обозначаящим  $\hat{X}(z_k)$ ) WDFТ может быть представлено выражением:

$$\begin{bmatrix} \hat{X}[0] \\ \hat{X}[1] \\ \vdots \\ \hat{X}[N-1] \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & A(z_0) & \cdots & A(z_0)^{N-1} \\ 1 & A(z_1) & \cdots & A(z_1)^{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & A(z_{N-1}) & \cdots & A(z_{N-1})^{N-1} \end{bmatrix}}_{\mathbf{D}} \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix}. \quad (3.4)$$

WDFТ наследует некоторые свойства ДПФ. Одним из важнейших является сопряженная симметрия для действительных данных:

$$\hat{X}[N-k] = \hat{X}^*[k]. \quad (3.5)$$

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

$$H_k(z) = \sum_{n=-\infty}^{\infty} A(z_k) z^{-n}, \quad k = 0 \dots N-1. \quad (3.6)$$

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

$$\hat{\omega}_k = 2 \tan^{-1} \left( \frac{1+a}{1-a} \tan \left( \frac{\omega_k}{2} \right) \right), \quad \omega_k = \text{angle}(z_k) \quad (3.7)$$

и полосой  $2\pi/N$ .

АЧХ фильтров  $H_k(z)$  для  $k = 0, \dots, N-1$  представлены на рис. 3.2.

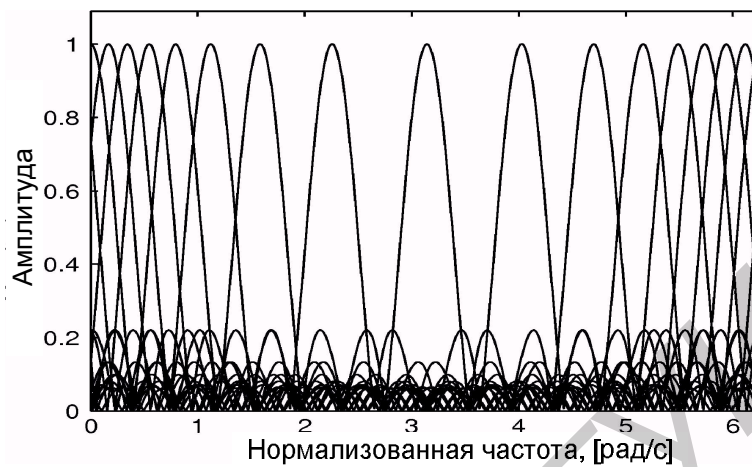


Рис. 3.2. АЧХ критически децимированного банка фильтров WDFT

Алгоритм, по эффективности сравнимый с БПФ, не может быть построен для WDFT из-за асимметричности WDFT-матрицы. Тем не менее, прямой алгоритм работы с комплексной матрицей может быть значительно оптимизирован. Наиболее усовершенствованный алгоритм, несмотря на сложность  $O(N^2)$ , использует факторизацию WDFT-матрицы в произведение действительной, ДПФ (осуществляемой через БПФ) и комплексной диагональной матрицы. Данный метод хорошо подходит для обработки изображений, где все данные поступают одновременно. В обработке аудио сигналов, где отсчеты поступают последовательно один за другим, может быть использована даже прямая реализация преобразования. Учитывая выражение (3.5), формула (3.4) может быть записана в следующем виде:

$$\begin{bmatrix} \hat{X}[0] \\ \hat{X}[1] \\ \vdots \\ \hat{X}[N-1] \end{bmatrix} = \sum_{n=0}^{N-1} \begin{bmatrix} A(z_0)^n \\ A(z_1)^n \\ \vdots \\ A(z_{N-1})^n \end{bmatrix} x[n]. \quad (3.8)$$

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

### 3.1.2. Обратное преобразование WDFT

В случае перцептуального деформирования оси частот матрицы Вандермонда, к классу которых относится и матрица  $\mathbf{D}$  WDFT, обладают свойствами, делающими невозможным получение точной обратной матрицы. А именно, матрица WDFT может рассматриваться как сингулярная так, как между некоторыми ее строками существуют почти линейные зависимости. Этот факт представляется как очень малое значение ее детерминанта  $\det \mathbf{D} = \prod_{i < j} (\hat{z}_i^{-1} - \hat{z}_j^{-1})$ .

Другими словами, WDFT-матрица является плохо обусловленной. Это означает, что любой численный алгоритм обращения матрицы, применяемый к подобной матрице является очень чувствительным к малым изменениям данных, которыми обычно пренебрегают. Усиление ошибки может быть оценено, используя число обусловленности матрицы (собственное значение матрицы)

$$\text{cond}(\mathbf{D}) = \|\mathbf{D}\| \|\mathbf{D}^{-1}\|, \quad (3.9)$$

где  $\|\cdot\|$  означает произвольную норму матрицы (в общем случае Евклидову). Чем больше число обусловленности, тем более некорректной становится задача.

Как показано на рис. 3.3, число обусловленности преобразуемой матрицы зависит от ее размерности и величины коэффициента деформации и имеет очень большие значения даже при малоразмерных преобразованиях, слегка отходящих от ДПФ. Плохая обусловленность является неотъемлемым свойством матриц Вандермонда, связанных с реальными проблемами. Единственным исключением является случай отсутствия деформирования частотной оси, когда WDFT становится обычным ДПФ.

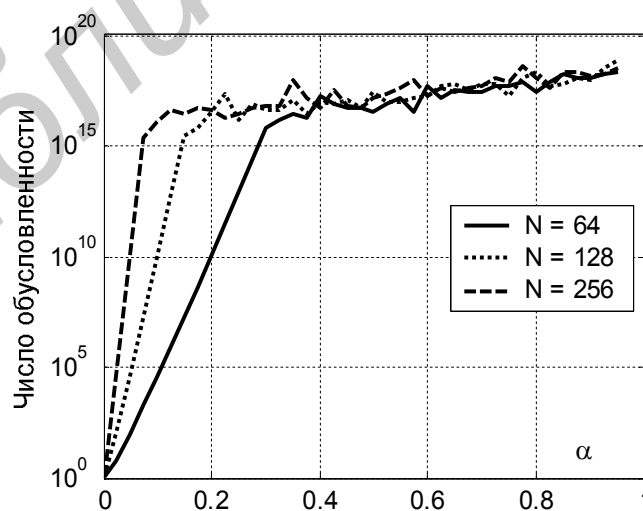


Рис. 3.3. Зависимость числа обусловленности от коэффициента деформации для различных размеров WDFT

Тем не менее, существует ряд методов для аппроксимации обращений плохообусловленных матриц. Данные методы используют разложение по сингулярным числам матрицы (от англ. – «Singular Value Decomposition» – SVD), которое определяется как

$$\mathbf{D} = \mathbf{U}\mathbf{Y}\mathbf{V}^H = \sum_{i=1}^N \mathbf{u}_i \sigma_i \mathbf{v}_i^H, \quad (3.10)$$

где

$$\begin{aligned} \mathbf{U} &= [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N] \\ \mathbf{V} &= [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N]^T \end{aligned} \quad (3.11)$$

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

$$\mathbf{Y} = \text{diag}(\sigma_1, \dots, \sigma_N) \quad (3.12)$$

является диагональной матрицей состоящей из сингулярных чисел, отсортированных по убыванию  $\sigma_1 \geq \dots \geq \sigma_N \geq 0$ . Столбцы матриц  $\mathbf{U}$  и  $\mathbf{V}$ , левый и правый, – сингулярные векторы.

Число обусловленности матрицы  $\mathbf{D}$  полного ранга есть отношение  $\text{cond}(\mathbf{D}) = \sigma_{\max} / \sigma_{\min}$ , где  $\sigma_{\max}$  и  $\sigma_{\min}$  – наибольшее и наименьшее сингулярные числа. Степень неполноты матрицы  $\mathbf{D}$  (наличие некоторых линейных зависимостей между ее столбцами) проявляется в существовании почти нулевых сингулярных чисел  $\sigma_i$ . Распределение сингулярных значений для нескольких различных матриц WDFТ проиллюстрированы на рис. 3.4. Как показывает рис. 3.4, значения достаточно большого количества сингулярных чисел близки к нулю.

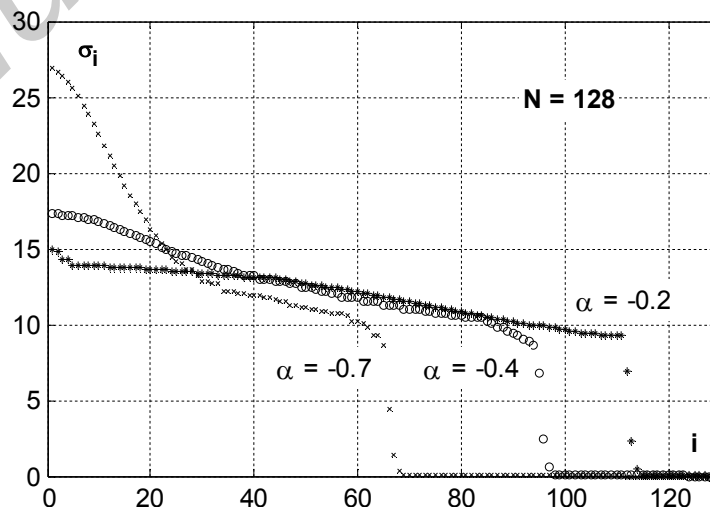


Рис. 3.4. Распределение сингулярных чисел как функция коэффициента деформации для фиксированного размера преобразования

Компоненты SVD могут дать много полезных пояснений плохой обусловленности матриц. Также, они могут быть использованы для формирования псевдообратной матрицы

$$\mathbf{D}^\dagger = \sum_{i=1}^N f_i \frac{1}{\sigma_i} \mathbf{v}_i \mathbf{u}_i^H \quad (3.13)$$

Символы  $f_i$  обозначают так называемые коэффициенты ослабления фильтра и должны быть все равны единице для получения точного обращения матрицы. Теория регуляризации рекомендует исключать влияние малых сигнулярных чисел, ослабляя их вклад в (3.13). Это реализуется путем установки соответствующих значений  $f_i$ . Так как сингулярные векторы с наибольшими индексами имеют ярко выраженный колебательный характер, то это действие изменяет спектральный состав данных.

Основное различие между известными методами регуляризации заключается в выборе коэффициентов ослабления фильтра. В простейшем подходе усеченного SVD (от англ. – «Truncated SVD» – TSVD) сумма (3.13) ограничивается обычным отбрасыванием термов имеющих сингулярные числа меньше определенного порога

$$f_i = \begin{cases} 1 & \sigma_i > \lambda, \\ 0 & \text{в противном случае.} \end{cases} \quad (3.14)$$

Менее радикальным является демпфированное SVD (от англ. – «Damped SVD» – DSVD), в котором коэффициенты ослабления фильтра постепенно изменяются в сторону нуля, обеспечивая более сглаженное отсечение.

$$f_i = \frac{\sigma_i}{\sigma_i + \lambda}. \quad (3.15)$$

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

### 3.1.3. Оценка ошибки реконструкции WDFT-синтеза

Используя матричную форму записи, вектор искажения сигнала можно определить как разность между оригинальным и реконструированным сигналами

$$\mathbf{d} = \mathbf{x} - \hat{\mathbf{x}} = (\mathbf{I} - \mathbf{D}^\dagger \mathbf{D}) \mathbf{x}. \quad (3.16)$$



Соответствующая мера в частотной области может быть выражена как средняя СПМ искажения сигнала

$$S_{dd}(\omega) = \frac{1}{N} E \left\{ \left| \mathbf{e}(\omega)^H \mathbf{d} \right|^2 \right\} = \frac{1}{N} \mathbf{e}(\omega)^H \mathbf{R}_{dd} \mathbf{e}(\omega)^H, \quad (3.17)$$

где

$$\mathbf{e}(\omega) = \left[ 1 \quad e^{-j\omega} \quad e^{-j2\omega} \quad \dots \quad e^{-j\omega(N-1)} \right]^T \quad (3.18)$$

является вектором ДПФ синусоид и  $\mathbf{R}_{dd}$  обозначает ковариационную матрицу сигнала искажения. Пусть  $\mathbf{Q} = \mathbf{I} - \mathbf{D}^\dagger \mathbf{D}$ , тогда

$$\mathbf{R}_{dd} = \mathbf{Q} \mathbf{R}_{xx} \mathbf{Q}^H, \quad (3.19)$$

где  $\mathbf{R}_{xx}$  ковариационная матрица обрабатываемого сигнала. Очевидно, что спектральное искажение (3.17), определяемое как абсолютная ошибка, зависит от характеристик входного сигнала и качества аппроксимации обратного WDFT. Теоретически, для ненулевого  $\mathbf{R}_{xx}$ , возможно точное обратное преобразование, когда все элементы равны нулю. Так как псевдообратная матрица рассчитывается методом разложения по сингулярным числам матрицы

$$\mathbf{Q} = \mathbf{I} - \mathbf{D}^\dagger \mathbf{D} = \mathbf{I} - \sum_{i=1}^N f_i \mathbf{v}_i \mathbf{v}_i^H, \quad (3.20)$$

подходящие настройки коэффициентов фильтра могут быть весьма полезны.

К сожалению, в случае перцептуальной деформации частотной оси, выбор параметров регуляризации имеет минимальное воздействие на уровень спектральных искажений. Было определено, что даже нестабильная аппроксимация обратного WDFT обеспечивает относительно высокую ошибку реконструкции и последняя усиливается при дальнейшей стабилизации.

Единственный путь минимизации спектрального искажения – это модификация матрицы преобразования таким образом, чтобы количество почти нулевых сингулярных чисел было уменьшено. Это идентично уменьшению эксцентricности SVD-эллипсоида, который является отображением единичной сферы в  $N$ -размерном пространстве.

На рис. 3.5 показана СПМ окрашенного гауссова шума и соответствующее искажение спектра, рассчитанное аналитически по (3.17). Можно заметить, что уровень искажений на заданной частоте зависит от расстояния между соседними WDFT коэффициентами. Сигнал полностью восстанавливается только в точках преобразования и спектральные искажения особенно заметны в растянутых частотных диапазонах, в то время как в сжатых диапазонах ошибка синтеза имеет приемлемый уровень.

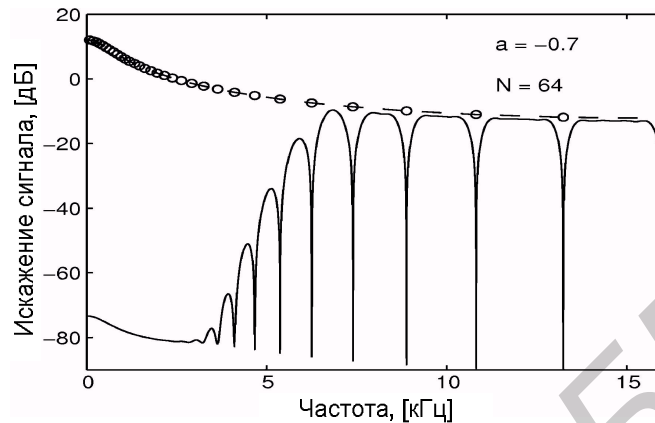


Рис. 3.5. СПМ оригинального сигнала и соответствующая ошибка синтеза

### 3.1.4. Формирование избыточного базиса WDFT

$k$ -я строка WDFT матрицы  $\mathbf{D}$  является комплексным синусоидальным вектором

$$\mathbf{s}_k = [1 \quad \hat{z}_k^{-1} \quad \hat{z}_k^{-2} \quad \dots \quad \hat{z}_k^{-N+1}]. \quad (3.21)$$

Теоретически, для конкретных точек преобразования строки WDFT матрицы линейно независимы (из-за ненулевого детерминанта матрицы Вандермонда). Поэтому, синусоидальные векторы WDFT (3.21) формируют не ортогональный базис  $\mathcal{S} = \{\mathbf{s}_0, \mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{N-1}\}$  для пространства комплексных векторов  $\mathbb{C}^N$ , так что любой входной вектор может быть выражен как их линейная комбинация.

Если мы ослабим ограничение на то, что векторы должны быть независимыми, то можно построить избыточный не ортогональный базис. Несмотря на то, что такой базис является избыточным, он может внести значительный вклад в коррекцию ошибки реконструкции, а именно, при подходящем выборе базиса векторов, мы можем модифицировать распределение сингулярных чисел для соответствующей матрицы преобразования.

Кроме того, так как мы предполагаем избыточный базис, новая WDFT матрица не является квадратной и количество строк  $M > N$  увеличивается. Матричное представление расширенного WDFT может быть записано следующим образом:

$$\begin{bmatrix} \hat{X}[0] \\ \hat{X}[1] \\ \vdots \\ \hat{X}[M-1] \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & \hat{z}_0^{-1} & \dots & \hat{z}_0^{-N+1} \\ 1 & \hat{z}_1^{-1} & \dots & \hat{z}_1^{-N+1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \hat{z}_{M-1}^{-1} & \dots & \hat{z}_{M-1}^{-N+1} \end{bmatrix}}_{\mathbf{D}_{M \times N}} \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix}, \quad (3.22)$$

Обобщенная инверсия прямоугольной матрицы может быть легко найдена, используя SVD-процедуру, аналогичную для квадратной матрицы.

С практической точки зрения, существуют как минимум два условия для построения избыточного векторного базиса WDFТ. Во-первых, необходимо сохранить регулярность точек преобразования (равномерное частотное разрешение в перцептуальной шкале). Во-вторых, размер нового синусоидального базиса должен быть по возможности минимальным для обеспечения минимизации вычислительной сложности.

Новый избыточный базис должен быть составлен из  $M > N$  векторов, соответствующих КИХ-фильтрам, чьи центральные частоты расположены регулярно по всей единичной окружности. Это может быть реализовано через переопределение точек преобразования:

$$z_k = e^{j\pi k/M}, \quad k = 0 \dots M-1. \quad (3.23)$$

Если предположить отрицательное значение параметра всепропускающего фильтра, то максимальное угловое расстояние между новыми точками  $z$ -преобразования составит

$$\Delta\omega_{\max} = \pi - 2 \tan^{-1} \left( \frac{1+a}{1-a} \tan \left( \frac{\pi - 2\pi/M}{2} \right) \right). \quad (3.24)$$

Для получения того же частотного разрешения в ВЧ-диапазоне, что и у обычного ДПФ, расстояние между точками преобразования не должно быть больше чем  $2\pi/N$ . Подставляя  $\Delta\omega_{\max} = 2\pi/N$  в (3.7) и решая относительно  $M$ , получается, что

$$M = M_{\text{opt}} = 2\pi \left[ \pi - 2 \tan^{-1} \left( \frac{1-a}{1+a} \tan \left( \frac{\pi - 2\pi/N}{2} \right) \right) \right]^{-1}. \quad (3.25)$$

Связь между параметрами  $a$  и  $M$  для различных размеров окна анализа представлено на рис. 3.6. Легко заметить, что для достаточно большого деформирования размер избыточного базиса быстро увеличивается. Рис. 3.7

показывает, что ошибка синтеза уменьшается при увеличении  $M$  и может не учитываться при  $M > M_{opt}$ .

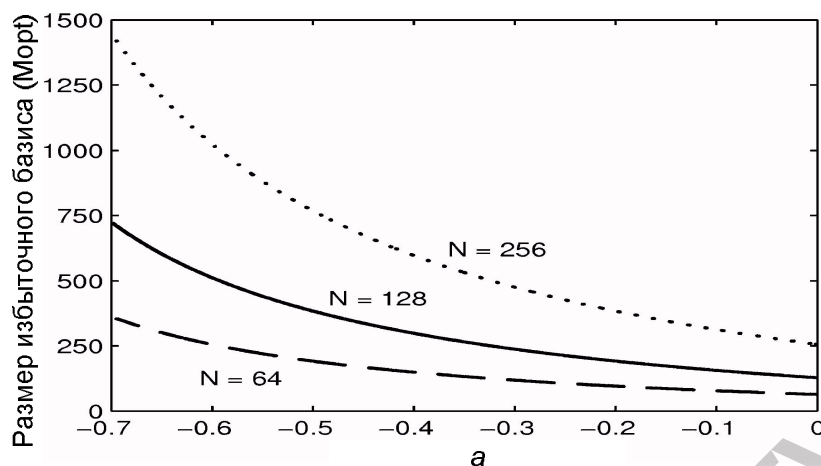


Рис. 3.6. Зависимость между значением параметра всепропускающего фильтра и размером избыточного базиса

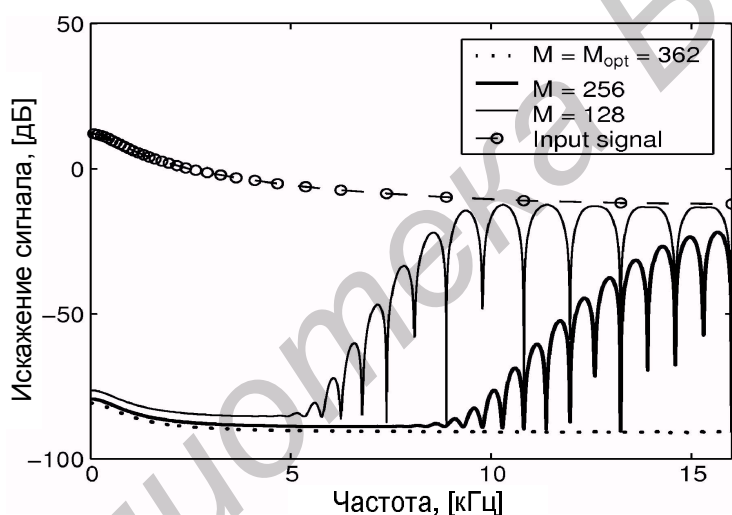


Рис. 3.7. СПМ входного сигнала и соответствующие спектральные искажения для разных размеров WDFT преобразования ( $N = 64$ ;  $M = 64, 100, 128$ )

### 3.1.5. Аппроксимация психоакустических шкал

Первый шаг при использовании WDFT в психоакустической модели – проектирование соответствующего всепропускающего преобразования. Частотные коэффициенты  $z$ - преобразования должны быть представлены равномерно в перцептуальной области. Показано, что всепропускающее звено первого порядка достаточно хорошо аппроксимирует перцептуальную шкалу Барков и ERB, при этом значение коэффициента всепропускающего фильтра для заданной частоты дискретизации определяется по выражениям (3.26) и (3.27), соответственно:

$$a_{Bark} = 0.1957 - 1.048 \cdot \left[ \frac{2}{\pi} \arctan \left( 0.07212 \frac{f_s}{1000} \right) \right]^{\frac{1}{2}}. \quad (3.26)$$

$$a_{ERB} = 0.7164 \cdot \left[ \frac{2}{\pi} \arctan \left( 0.09669 \frac{f_s}{1000} \right) \right]^{\frac{1}{2}} + 0.08667. \quad (3.27)$$

Для случая  $f_s = 16\text{kHz}$  коэффициент  $a_{Bark} = -0.57827$ . На рис. 3.8 представлена трансформация (деформирование) частотной оси согласно шкале Барков, ERB и MEL для  $f_s = 16\text{kHz}$ .

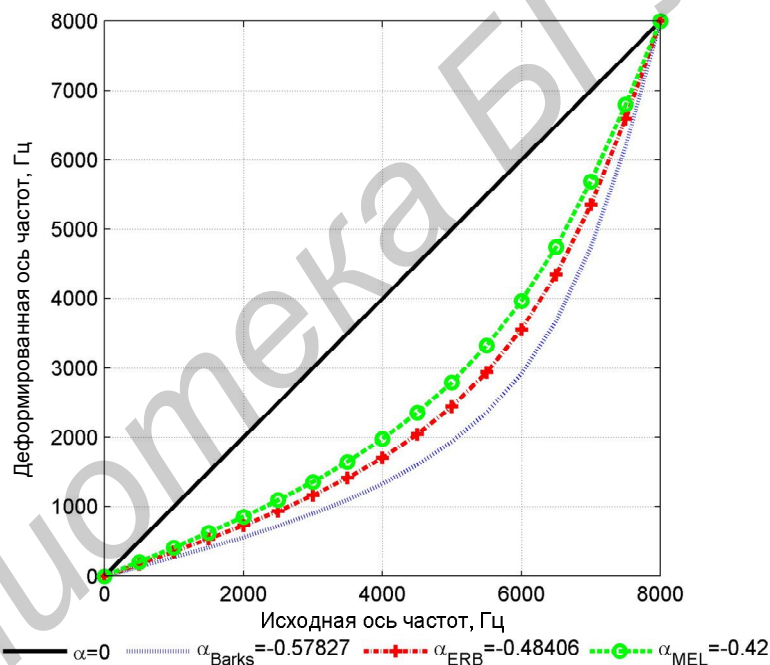


Рис. 3.8. Деформация частотной оси в соответствии с перцептуальными шкалами

Так как ДПФ имеет равномерное частотное разрешение, в то время как ширина критических частотных полос строго изменяется с их местоположением на частотной шкале, то различное количество коэффициентов преобразования ассоциируется с конкретной критической частотной полосой. В части А табл. 3.1 количество коэффициентов в группах варьируется от 3 до 38, в то время как для WDFТ той же размерности не отдается предпочтения ни одной из полос, все коэффициенты преобразования распределены практически равномерно (часть Б).

Таблица 3.1.

Отображение коэффициентов ДПФ и WDFТ на критические частотные полосы (Барки)

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

WDFТ в его оригинальной форме не сохраняет энергию сигнала в соответствующих частях единичной окружности до и после всепропускающего преобразования (рис. 3.9). Так как искривление приводит к растяжению одного частотного диапазона при одновременном сжатии другого, оно должно сопровождаться масштабированием WDFТ-амплитуд с целью коррекции уровней энергии в каждой критической частотной полосе. Предлагается следующая коррекция:

$$\hat{X}_{PC}(z) = \frac{\sqrt{1-a^2}}{1-az} \hat{X}(z). \quad (3.28)$$

Рассмотрим в качестве примера широкополосный аудио сигнал, который содержит голосовое сообщение (первые 3.5с) и музыкальный фрагмент с вокальным сопровождением (рис. 3.10(сверху)). Построим спектрограммы данного сигнала при помощи ДПФ и WDFТ (рис. 3.10 центр и низ, соответственно). Как видно из рис. 3.10 WDFТ-решение дает достаточно четкий результат, оценка энергии более точная. Представленные факты свидетельствуют о том, что WDFТ может действительно служить базисом для психоакустической модели.

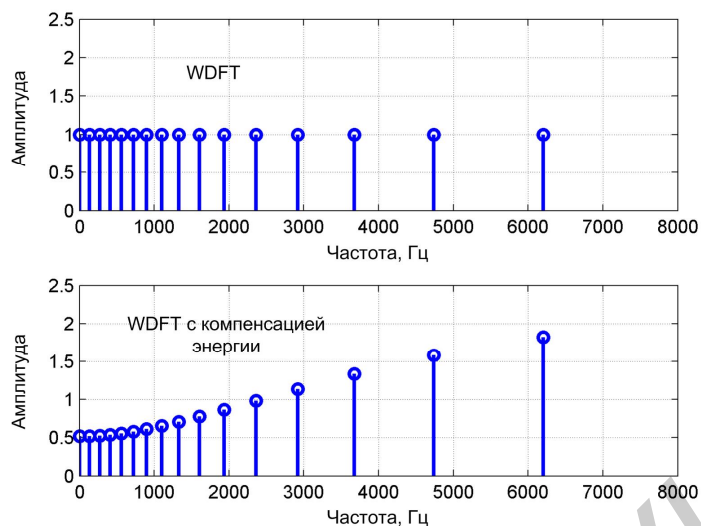


Рис. 3.9. Компенсация энергии WDFT ( $a_{Bark} = -0.57827$ ) единичного импульса

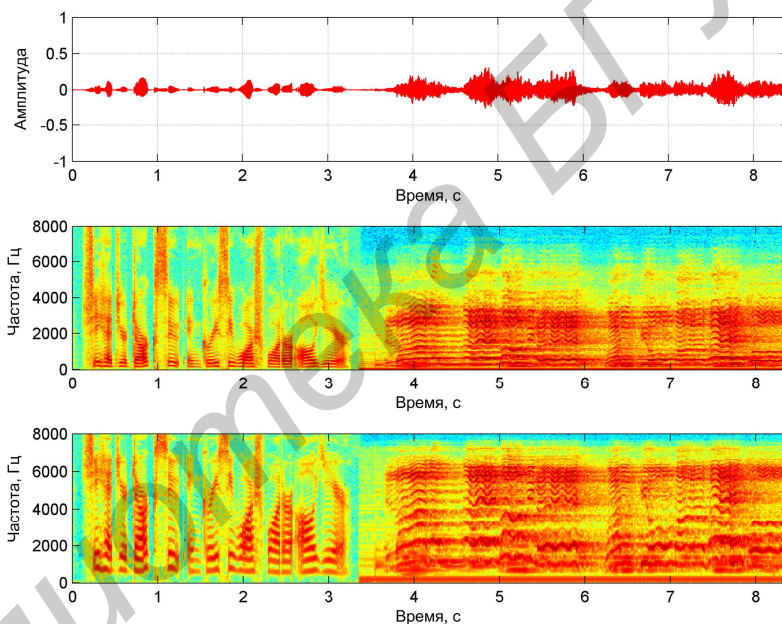


Рис. 3.10. Широкополосный сигнал во временной области (сверху) и его спектрограммы, полученные при помощи ДПФ (в центре) и WDFT с коэффициентом  $a_{Bark} = -0.57827$  (снизу)

Ниже представлен листинг MATLAB®-программы расчета прямой и обратной матрицы WDFT.

```
function [W, W_1] = wdfmtmx(N,M,a)
% формирование WDFT-матрицы размера NxM с коэффициентом деформации a
% [ e^(-jw[1]) e^(-jw[1]) ... e^(-jw[1]) ] .^ [ 0 1 ... M-1 ]
% [ e^(-jw[2]) e^(-jw[2]) ... e^(-jw[2]) ] [ 0 1 ... M-1 ]
% [ ... .. ] [ ... .. ]
% [ e^(-jw[N]) e^(-jw[N]) ... e^(-jw[N]) ] [ 0 1 ... M-1 ]
N2 = ceil(N/2);
z_1 = exp(-j * 2 * pi * (0 : (N2 - 1)) ./ N) .';
z_1 = (z_1 - a) ./ (1 - a .* z_1);
W = (z_1*ones(1,M)).^(ones(N2,1)*[0:M-1]);
```

```

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

if nargin < 2, return; end;

% формирование псевдоинверсной матрицы по методу TRUNCATED SVD
% W = U*S*V', где U*U' = I и V*V' = I
%
%  $\bar{W}_1 = V \bar{S} U'$ 
%
%  $\bar{S} = \text{diag}( [ s[1]^{-1} s[2]^{-1} \dots s[K]^{-1} 0 \dots 0 ] )$ , где  $s[K] > \lambda$ 

[U, S, V] = svd(W,0);
s = diag(S);
% plot(s); pause; %сними комментарий, чтобы увидеть распределение сингулярных
% чисел матрицы

lambda = max(s)/10; % порог
S = diag([s(find(s)>lambda)].^-1; 0*s(find(s)<=lambda)]);
 $\bar{W}_1 = V \bar{S} U'$ ;

```

## 4. Применение банков цифровых фильтров и WDFT в системах мультимедиа

### 4.1. Перцептуально мотивированный широкополосный кодер речи

#### 4.1.1. Структура кодера и декодера

Предлагается следующая структурная схема широкополосного CELP-кодера (рис. 4.1 и рис. 4.2), ядром которой является мультиполосная кодовая книга, состоящая из субполосных кодовых книг форм и коэффициентов усиления. Входной сигнал квантуется при помощи 16 бит/отсчет с частотой дискретизации  $f_s = 16 \text{ кГц}$ , после чего кодируется широкополосным кодером, структура которого представлена на рис. 4.1. Кодер использует блочный LP-анализ, без интерполяции коэффициентов линейного предсказания. В качестве наиболее оптимального для полосы частот 50 – 8000 Гц, был определен  $p = 16$  порядок фильтра-предсказателя (Short-Term Predictor - STP). Для улучшения работоспособности взвешивающего фильтра (осуществления независимого контроля формы спектральной огибающей и формантного взвешивания) используется предсказывающий фильтр.

Кодер имеет восьмиуровневую структуру, т.е. количество уровней соответствует количеству обрабатываемых субполос. Поиск осуществляется последовательно по мере перцептуальной важности полос, т.е. от низкочастотных к высокочастотным. Очередной оптимальный вектор с соответствующим коэффициентом усиления пропускается через последовательно соединенные LTP- и взвешивающий синтезирующий фильтр.



По окончании цикла минимизации ошибки из остаточного взвешенного оригинального сигнала вычитается вклад очередной субполосы (уровня). Процесс выполняется последовательно для всех восьми субполос. После обработки всех полос, производится синтез речевого вектора по оптимальному восьмиполосному возбуждающему сигналу, LTP-буфер обновляется синтезированным вектором сигнала, а также осуществляется обновление памяти взвешивающего синтезирующего фильтра. Алгоритм повторяется для оставшихся субфреймов.

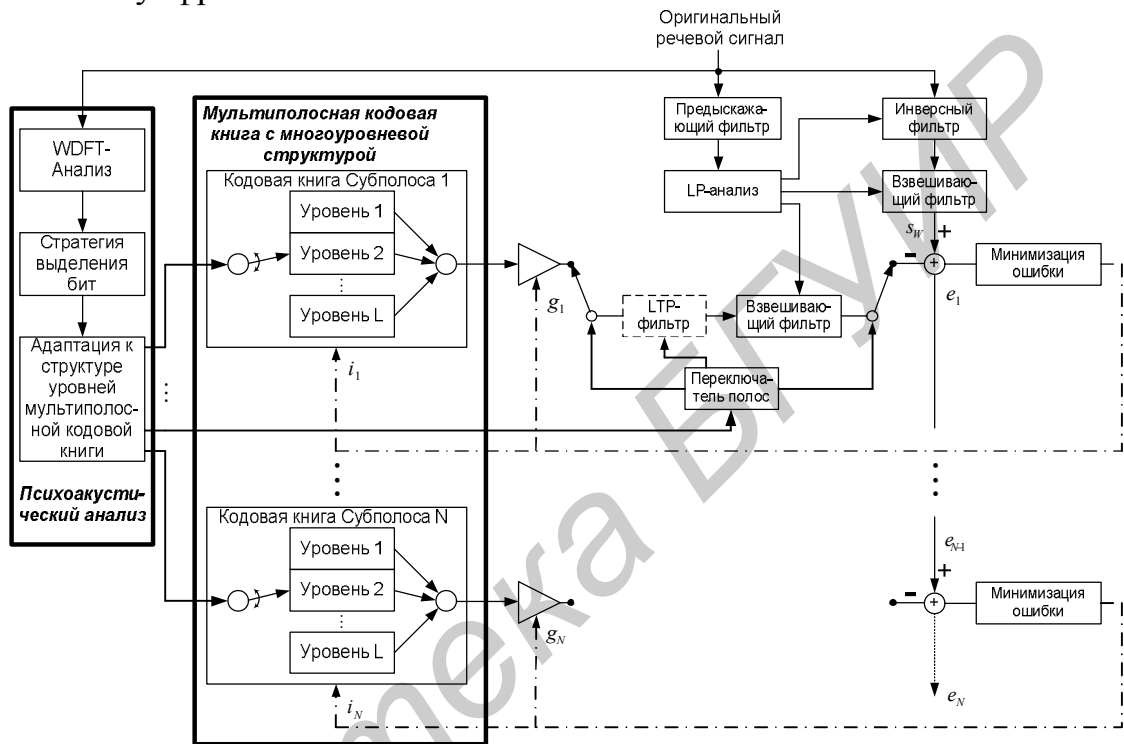


Рис. 4.1. Структура широкополосного кодера

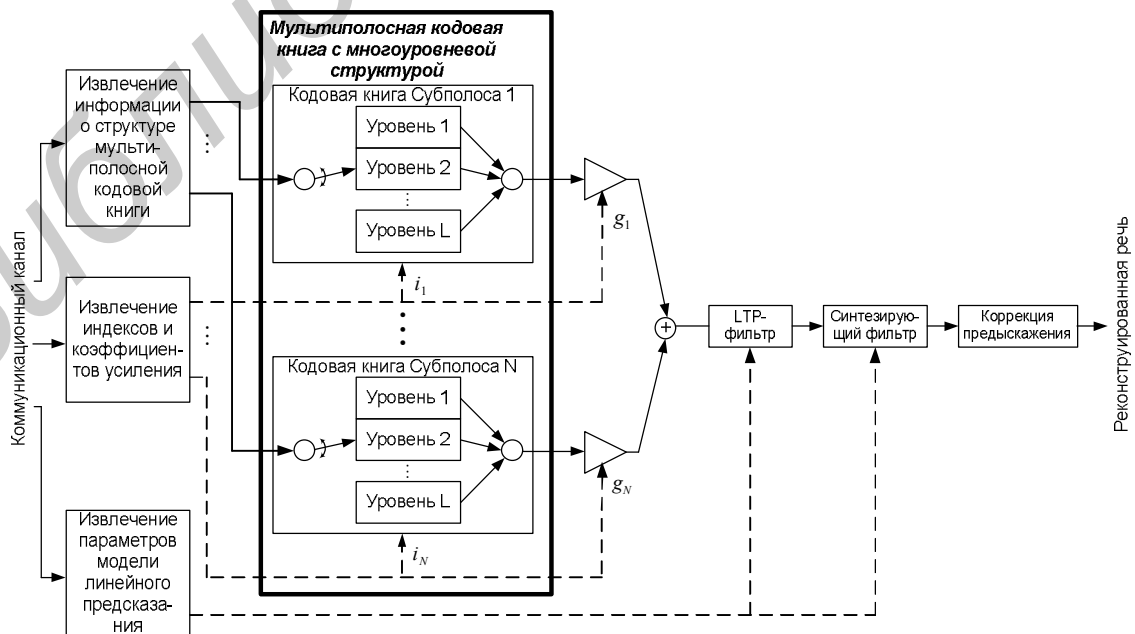


Рис. 4.2. Структура широкополосного декодера

#### 4.1.2. Разбиение частотной оси на субполосы

В рассматриваемом широкополосном перцептуальном CELP-кодере с целью отражения частотно зависимого поведения аудиторной системы человека принята шкала барков и используется схема разбиения на субполосы, представленная в табл. 4.1.

Выбор субполос – результат информативных тестов прослушивания, а также оптимизации критериев описанных в п.2. Ширина полос увеличивается с частотой с целью отражения частотной избирательности аудиторной системы человека и формантной структуры речи. Для формирования субполосных обучающих множеств был использован полифазный банк фильтров с косинусной модуляцией и неравномерным разбиением на полосы, согласно табл. 4.1. АЧХ банка фильтров анализа представлена на рис. 4.3.

Таблица 4.1.

Характеристика субполос кодера

Субполоса	Частотный диапазон, Гц	Барки
1	100-510	4
2	510-1080	4
3	1080-1720	3
4	1720-2320	2
5	2320-3150	2
6	3150-4100	1.5
7	4100-5300	1.5
8	5300-8000	3

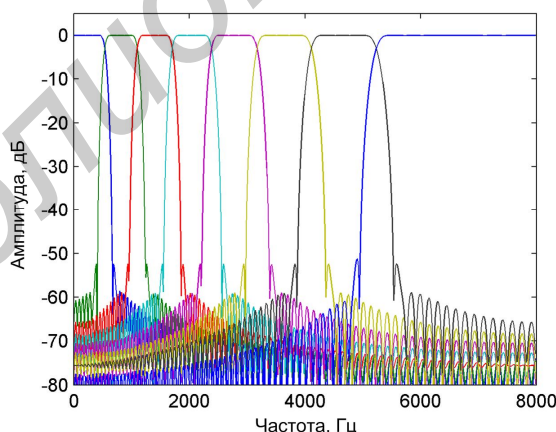


Рис. 4.3. АЧХ банка фильтров

Необходимо отметить, что отфильтрованные кодовые векторы имеют конечную длительность (равную длине субфрейма), вследствие чего существует утечка «просачивание» энергии между смежными кодовыми книгами (субполосами). Однако, так как кодовые векторы разных кодовых книг

близко ортогональные, последовательный поиск в каждой субполосной кодовой книге обеспечивает почти тоже качество, что оптимальный совместный поиск во всех субполосных книгах, но со значительным снижением сложности. Более того, «просачивание» спектральной энергии  $i$ -й субполосы может быть компенсировано при квантовании  $i+1$ -й субполосы, поэтому нет необходимости в банке фильтров с перфективной реконструкцией.

#### 4.1.3. Реконфигурирование структуры кодовой книги на основе субполосной перцептуальной энтропии

Оценка порога маскирования основана на общей психоакустической модели, предложенной Джонстоном в 1988 году. Данная процедура была подстроена под широкополосную речь с полосой 8 кГц и модифицирована для использования свойств WDFT. Основные шаги оценки порога маскирования следующие: расчет энергии критических частотных полос по WDFT-спектру мощности; свертка с функцией распространения по базилярной мембране; вычитание относительного смещения порога; нормализация и сравнение с абсолютным порогом слышимости.

Энергия рассчитывается в критических частотных полосах путем суммирования компонент спектра мощности в каждой критической частотной полосе. Коэффициенты преобразования группируются в субполосы в соответствии с табл. 3.1 и табл. 4.1, в каждой из которых осуществляется оценка порогов маскирования. Как упоминалось выше, в случае WDFT мы имеем неравномерную концентрацию энергии в окрестностях преобразуемых центральных частот. Поэтому, для вычисления энергии, сосредоточенной в критических частотных полосах, необходимо учесть масштабирование (3.28).

По вычисленной энергии порога маскирования осуществляется определение перцептуальной важности кодируемых субполос или субполосная перцептуальная энтропия:

$$SPE_b = \log 2 \left( \frac{E_b}{\sqrt{\frac{6 \cdot T_b}{\text{bandwidth}_b}}} + 1 \right), \text{ [бит/полосу]} \quad (4.1)$$

где  $E_b$  - энергия субполосы,  $T_b$  - субполосный порог маскирования,  $\text{bandwidth}_b$  - ширина субполосы  $b$  в спектральных отсчетах.

Для того чтобы преобразовать перцептуальную энтропию каждой субполосы  $SPE_b$  в глубину поиска в соответствующей субполосной кодовой книге  $\text{depth}_b$ , используется масштабирование вектора  $SPE$  для каждого анализируемого фрейма входного сигнала таким образом, чтобы максимальное

значение в векторе  $SPE_{\max}$  соответствовало максимально возможной глубине поиска в субполосной кодовой книге  $depth_{b,MAX}$ . Для этого вычисляется вклад каждого уровня субполосной книги  $PE\_per\_stage_b$  в суммарное значение субполосной перцептуальной энтропии  $SPE_{\max}$ :

$$\begin{aligned}
 & \text{Если } depth_{b,MAX} \geq \max_{b \in B}(SPE_b) \\
 & PE\_per\_stage_b = \frac{\max_{b \in B}(SPE_b)}{depth_{b,MAX}}, \\
 & \text{иначе} \qquad \qquad \qquad \text{для } b = 1, 2, \dots, 8 \quad (4.2) \\
 & PE\_per\_stage_b = \frac{depth_{b,MAX}}{\max_{b \in B}(SPE_b)}; \\
 & depth_b = round(SPE_b \cdot PE\_per\_stage_b)
 \end{aligned}$$

где  $B$  - множество субполос,  $round$  - операция округления.

Принцип изменения структуры книги заключается в следующем. Оригинальный речевой сигнал (очередной кодируемый фрейм) поступает на вход блока WDFТ-анализа, в котором осуществляется оценка субполосных порогов маскирования  $T_b$ , затем производится оценка перцептуальной энтропии каждой из субполос кодера  $SPE_b$ . Таким образом, на выходе мы имеем своего рода стратегию выделения бит или вектор  $SPE$ , содержащий значения перцептуальной важности кодируемых субполос для текущего фрейма сигнала. Данная стратегия, согласно алгоритму (4.2), преобразуется в вектор  $depth$ , содержащий глубину поиска (номер уровня детализации векторов возбуждения) в субполосных кодовых книгах. При этом если  $SPE_b = 0$ , то данная субполоса удаляется из рассмотрения, т.е. не кодируется индекс вектора возбуждения и его коэффициент усиления. Вектор  $depth$  квантуется по 10-битной кодовой книге и передается на сторону декодера. Если обозначить  $Stage\_Size_b[i]$  размер  $i$ -го уровня субполосной кодовой книги  $b$  в векторах, то количество бит, необходимых для кодирования индексов векторов возбуждения в многоуровневой мультиполосной кодовой книге, составит

$$\text{Book Indexes} = 4 \cdot \sum_{\substack{b=1 \\ SPE_b > 0}}^8 \log_2(Stage\_Size_b[depth_b]), \text{ [бит/фрейм]}, \quad (4.3)$$

а для коэффициентов усиления

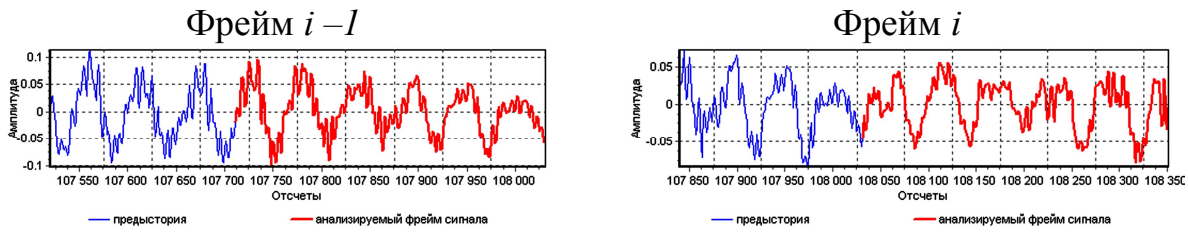
$$\text{Excitation Gains} = 4 \cdot \sum_{\substack{b=1 \\ SPE_b > 0}}^8 4, [\text{бит/фрейм}]. \quad (4.4)$$

Таким образом, схема квантования параметров модели речеобразования преобразуется к виду, представленному в табл. 4.2.

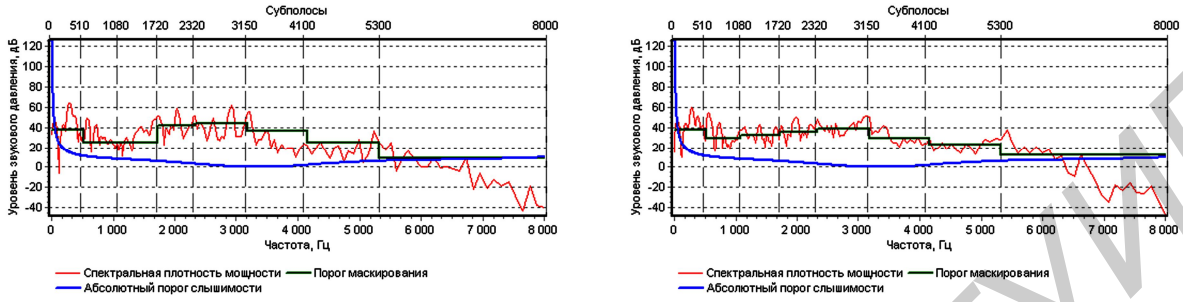
Таблица 4.2.  
Характеристика квантования параметров модели речеобразования широкополосного CELP-кодера с реконфигурируемой структурой кодовой книги

Параметр модели	Кол-во парам./фрейм	Кол-во бит/параметр	Суммарное количество бит/фрейм	Кол-во фреймов/с	Скорость потока, бит/с
LSF	16	1.6875	27	50	1350
Model Gain	1	7	7		350
LTP Delays	4	8	32		1600
LTP Gains	4	5	20		1000
<i>depth</i>	1 вектор из 8 элементов	1.25	10		500
Excitation Gains	Выражение (4.4)	4	VAR		<b>VAR</b>
Book Indexes	Выражение (4.3)	VAR	VAR		<b>VAR</b>
<b>Результирующий поток данных:</b>					<b>VAR</b>

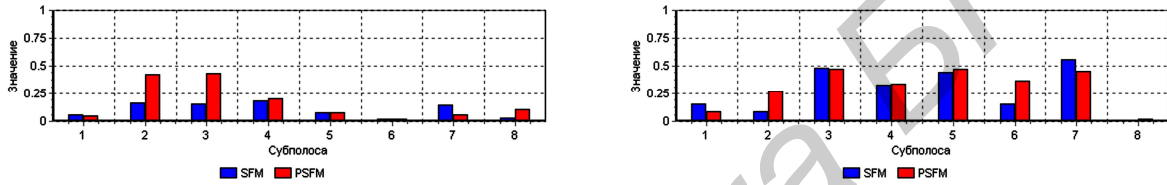
Процесс реконфигурации мультиполосной многоуровневой кодовой книги по этапам представлен на рис. 4.4. Буферизация анализируемого речевого сигнала (рис. 4.4а) позволяет учитывать свойства предыдущего анализируемого фрейма, при этом можно избежать резких изменений в структуре кодовой книги (рис. 4.4е). На рис. 4.4б проиллюстрированы пороги маскирования в субполосах для двух последовательных фреймов. Анализ рис. 4.4в,г,д, показывает, что чем меньше значение меры пологости спектра субполосного сигнала, тем больше величина отношения сигнал-маска и, следовательно, значение субполосной перцептуальной энтропии. На рис.4.5 приведен пример изменения скорости потока в зависимости от перцептуальной важности сигнала.



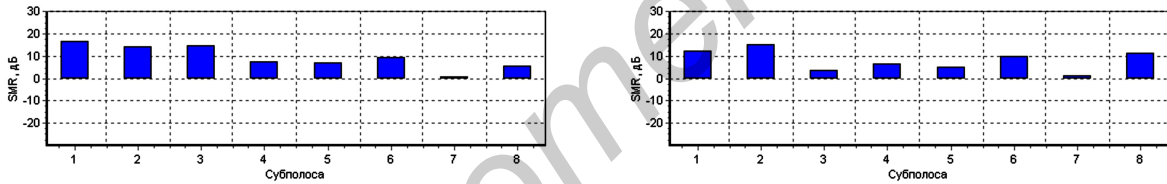
а) буфер анализируемого речевого сигнала



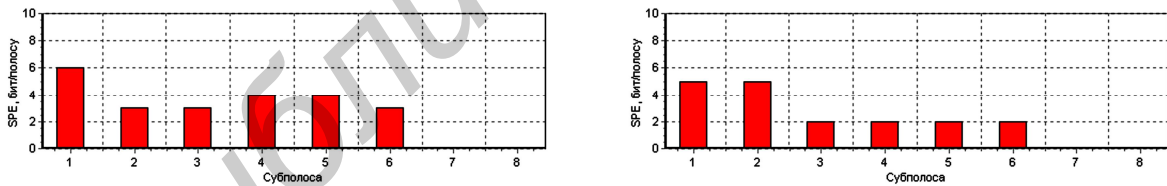
б) пороги маскирования в субполосах



в) оценка меры пологости спектра *SFM* и *PSFM* в субполосах



г) соотношение сигнал-маска *SMR* в субполосах



д) величины субполосной перцептуальной энтропии



е) структура мультиполосной многоуровневой кодовой книги

Рис. 4.4. Пример работы психоакустической модели с результатами реконфигурации мультиполосной многоуровневой кодовой книги

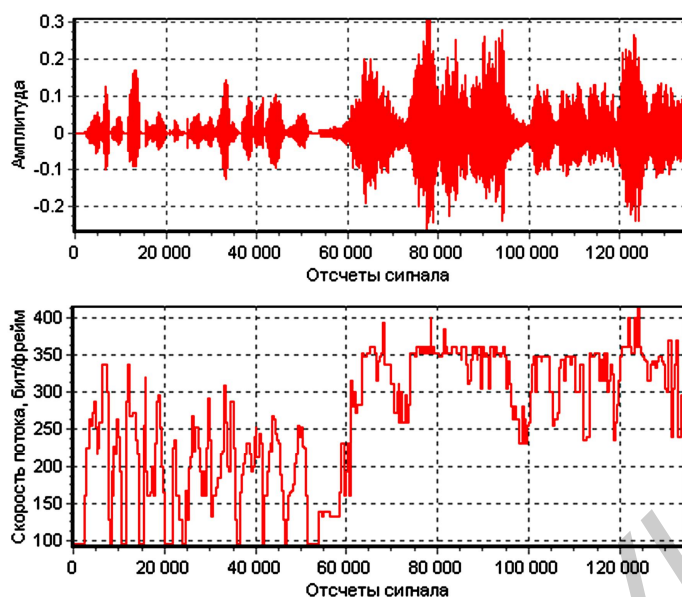


Рис. 4.5. Изменение скорости потока в зависимости от перцептуальной важности сигнала: огибающая сигнала (сверху), поток данных (снизу)

Как видно из рис. 4.5, предлагаемая схема психоакустической оптимизации (реконфигурации) структуры мультиполосной многоуровневой кодовой книги кодера позволяет существенно снизить вклад данных об индексах векторов возбуждения и коэффициентах их усиления в результирующий информационный поток. При этом увеличение количества бит, необходимых для кодирования информации, наблюдается лишь в перцептуально значимых областях (фрагментах) речевого сигнала.

## 4.2. Система редактирования шума в речевом сигнале

### 4.2.1. Введение

Большинство существующих систем подавления шума работают в частотной области, при этом используется хорошо известный подход *спектрального вычитания*. Несмотря на то, что эти методы очень просты и легко реализуются, их слабым местом является остаточный шум, также известный как «музыкальный тон». Следовательно, возникает задача модификации взвешивающего правила таким образом, чтобы оставить музыкальные тона немного ниже порога маскирования.

Применение WDFT в перцептуальной системе редактирования шума не только как базиса для определения модели маскирования, но и как инструмента спектрального анализа, позволяет добиться наилучшего качества по сравнению с обычными системами повышения качества речи. Это объясняется тем, что весь процесс обработки осуществляется в перцептуальном домене с неравномерной частотной шкалой (от англ. – «Warped Spectrum Domain») и нет необходимости в преобразованиях между разными частотными шкалами, что приводит к упрощению архитектуры системы. Более того, обработка речи,

осуществляемая в критических частотных полосах, более точна в контексте психоакустического моделирования.

#### 4.2.2. Краткое описание системы

Структурная схема препроцессора повышения качества речи на базе WDFT представлена на рис. 4.6. Алгоритм основан на хорошо известном правиле спектрального взвешивания.

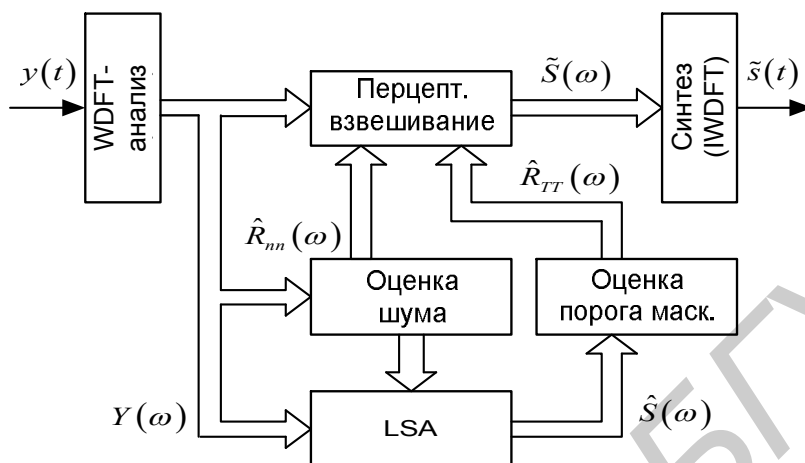


Рис. 4.6. Структура перцептуальной системы подавления шума на базе WDFT

Предполагается, что чистый речевой сигнал  $s(i)$  и окружающий шум  $n(i)$  статистически независимы и стационарны в широком смысле ( $i$  обозначает временной индекс). Тогда оцифрованный зашумленный речевой сигнал может быть представлен следующим образом:

$$y(i) = s(i) + n(i). \quad (4.5)$$

Обработка осуществляется в частотной области последовательными фреймами. Входной сигнал  $y(i)$  разбивается на перекрывающиеся фреймы из  $N$  последовательных отсчетов. Каждый фрейм умножается на оконную функцию и преобразуется в частотную область. Результат обозначим как  $Y^{(\lambda)}(\omega)$ , где  $0 < \omega < 2\pi$  и  $\lambda$  - индекс фрейма (далее для упрощения будем его опускать). Уменьшение шума достигается умножением спектральных коэффициентов зашумленной речи на действительные коэффициенты взвешивающей функции  $H^{(\lambda)}(\omega)$

$$\tilde{S}(k) = H(\omega) \cdot Y(\omega), \quad 0 \leq H(\omega) \leq 1. \quad (4.6)$$

Заметим, что фаза сигнала не изменяется. После модификации амплитуд, спектр очищенного сигнала  $\tilde{S}(\omega)$  преобразуется обратно во



временную область и синтезируется при помощи метода перекрытия с накоплением.

#### 4.2.3. Психоакустически мотивированное правило спектрального взвешивания

Основной задачей данного подхода является не полное удаление шума, а ослабление слышимого шума на постоянный коэффициент. Это обусловлено тем, что в некоторых случаях, например, во время диалога по телефону, полное удаление шума не желательно. Для того чтобы сохранить характеристики окружающего шума необходимо определить предпочтительный уровень остаточного шума  $\zeta_n$ . Тогда разница между желаемым спектром очищенной речи и его оценкой может быть определена как

$$Q(\omega) = S(\omega) + \zeta_n N(\omega) - H(\omega)[S(\omega) + N(\omega)], \quad (4.7)$$

где  $S(\omega)$  и  $N(\omega)$  спектры чистой речи и шума, соответственно, и  $H(\omega)$  – оцененная взвешивающая функция.

Так как речь и шум предполагаются статистически независимыми, то СПМ разности (4.7) может быть выражена следующим образом:

$$R_{qq}(\omega) = [1 - H(\omega)]^2 R_{ss}(\omega) + [\zeta_n - H(\omega)]^2 R_{nn}(\omega) \quad (4.8)$$

или

$$R_{qq}(\omega) = R_{qsqs}(\omega) + R_{qnqn}(\omega). \quad (4.9)$$

Слагаемые выражения (4.9) соответствуют СПМ искажения речи и шума, соответственно. Для минимизации суммы (4.9) в перцептуальном смысле необходимо чтобы уровень искажений был невоспринимаемым на слух. В идеальном случае все искажения должны быть замаскированы. Тем не менее, в большинстве реальных систем это требование не может быть удовлетворено, так как  $R_{qq}(\omega)$  может быть больше чем порог маскирования. Поэтому критерий минимизации формулируется как

$$R_{qnqn}(\omega) = R_{TT}(\omega). \quad (4.10)$$

Решая уравнение (4.8) относительно  $H(\omega)$ , приходим к взвешивающему правилу

$$H^{IND}(\omega) = \min \left\{ 1, \sqrt{\frac{R_{TT}(\omega)}{R_{nn}(\omega)} + \zeta_n} \right\}, \quad (4.11)$$

где  $IND$  означает неслышимое шумовое искажение.

Легко заметить, что если остаточный шум лежит ниже порога маскирования, выражение под корнем больше единицы и речь не искажается так как  $H^{IND}(\omega) = 1$ . В противном случае, окружающий шум оптимально ослабляется до невоспринимаемого уровня. Для расчета взвешивающих коэффициентов необходима оценка порога маскирования и СПМ шума.

#### 4.2.4. Оценка СПМ шума

Оценка СПМ шума достаточно сложная задача и в различных системах повышения качества речи может решаться по-разному. В рассматриваемой системе (рис. 4.6) оценку шума предлагается реализовать по методу экспоненциального усреднения с контролем минимума энергии (от англ. – «Minima-Controlled Recursive Averaging» – MCRA). Данное решение является промежуточным между обычным измерением энергии шума с рекурсивным усреднением и методом минимума статистики. Кроме того, данный метод реализуется в перцептуальной области.

Предполагается, что СПМ шума в каждой критической частотной полосе постоянна (данное утверждение верно в большинстве практических случаев)

$$R_{nn}(\omega) = R_{nmb}, \quad \omega_{lb} < \omega < \omega_{hb}, \quad (4.12)$$

где  $\omega_{lb}$ ,  $\omega_{hb}$  - границы критической частотной полосы и  $b$  - Барк.

СПМ шума в каждой критической частотной полосе вычисляется как

$$\hat{R}_{nmb}^{(\lambda)} = \alpha_b^{(\lambda)} \hat{R}_{nmb}^{(\lambda-1)} + (1 - \alpha_b^{(\lambda)}) \hat{R}_{yub}^{(\lambda)}, \quad (4.13)$$

где  $\hat{R}_{yub}^{(\lambda)}$  - СПМ зашумленного сигнала, оцененная на текущем фрейме и  $\alpha_b^{(\lambda)}$  - изменяемый во времени параметр сглаживания, настраиваемый на основе вероятности вокализованности  $p_b^{(\lambda)}$   $\lambda$ -го фрейма следующим образом:

$$(4.14)$$

$$\alpha_b^{(\lambda)} = \alpha_{slow} + (1 - \alpha_{slow})p_b^{(\lambda)}, \quad 0 < \alpha_{slow} < 1.$$

Параметр  $\alpha_{slow}$  - коэффициент сглаживания, который контролирует общую способность слежения при оценке энергии шума. Вероятность вокализованности фрейма  $p_b^{(\lambda)}$  рассчитывается на основе экспоненциального усреднения решений  $I_b^{(\lambda)}$  простого детектора речевой активности (от англ. – «Voice Activity Detector» – VAD), основанного на энергии фрейма

$$p_b^{(\lambda)} = \alpha_{fast}p_b^{(\lambda-1)} + (1 - \alpha_{fast})I_b^{(\lambda)}, \quad 0 < \alpha_{fast} < 1, \quad (4.15)$$

где  $\alpha_{fast}$  - параметр сглаживания.

Выход VAD  $I_b^{(\lambda)}$  может быть интерпретирован как логический результат сравнения смещенной оценки минимума статистики СПМ шума и зашумленного сигнала (если оценка СПМ зашумленного сигнала в два раза больше оценки минимума статистики шума генерируется «1»).

Дополнительное частотное усреднение среди критических частотных полос уменьшает дисперсию оценки измеренной СПМ шума. Несмотря на то, что частотное разрешение также снижается, это не является большой проблемой в случае оценки шума, так как большинство шумов имеют простые спектральные характеристики. Более того, разрешение снижается неравномерно в соответствии с частотной избирательностью человеческого уха. Для обычного ДПФ усреднение в критических частотных полосах приводит к расчету арифметического среднего энергии всех спектральных компонент критической частотной полосы. Для WDFT частотное усреднение может быть обобщено следующим образом:

$$\hat{R}_{yyb} = \frac{1}{\omega_{hb} - \omega_{lb}} \sum_{k \in K_b} \Delta\omega_k |Y(\omega_k)|^2, \quad (4.16)$$

где  $Y(\omega_k)$  - WDFT (или ДПФ) спектр,  $K_b$  - множество частотных отсчетов, принадлежащих критической частотной полосе (например, см. табл. 3.1) и  $\Delta\omega_k$  - ширина полосы соответствующего частотного отсчета.

По уравнению (4.16) осуществляется интегрирование СПМ-функции по всем неравномерно расположенным узлам. При этом для обычного ДПФ (4.16) может быть упрощено и представлено в следующем виде:

$$\hat{R}_{yyb}^{ДПФ} = \frac{1}{C(K_b)} \sum_{k \in K_b} |Y(\omega_k)|^2, \quad (4.17)$$

где  $C(K_b)$  - мощность множества  $K_b$ .

Определение порога маскирования чистой речи основано на модифицированной модели перцептуальной энтропии Джонстона. Для удаления «музыкального тона» в данной системе используется правило логарифмов спектральных амплитуд (от англ. – «Log Spectral Amplitude» – LSA) и метод управления принятием решения.

#### 4.2.5. Оценка качества системы подавления шума

На рис. 4.7 представлен сравнительный анализ качества системы подавления шума, реализованной на базе ДПФ, WDFT и WDFT с избыточным базисом. Спектрограммы сигналов показаны на рис. 4.8.

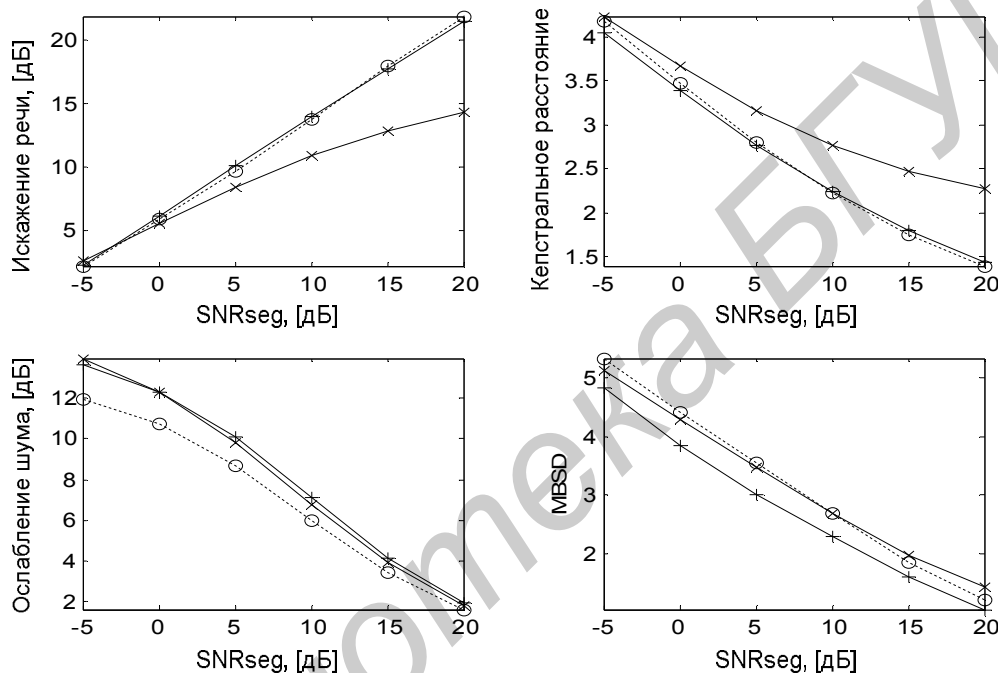


Рис. 4.7. Оценка качества систем на базе ДПФ (+), WDFT (x), WDFT с избыточным базисом (o)

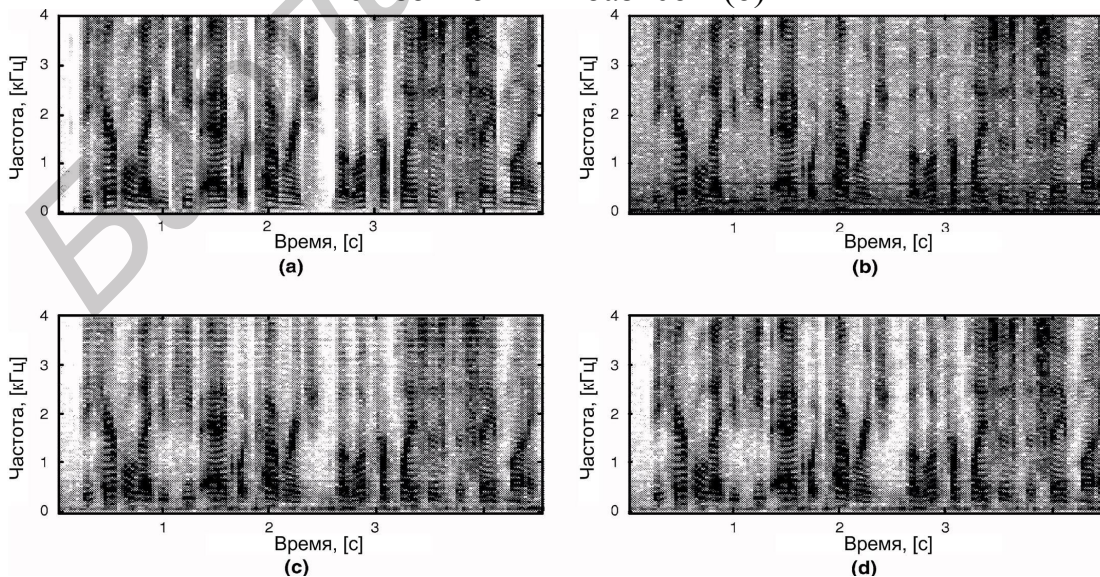


Рис. 4.8. Спектрограммы: оригинальный сигнал (а); зашумленный речевой сигнал SNRseg=5дБ (b); сигнал, очищенный системой на базе WDFT (c); сигнал, очищенный системой на базе WDFT с избыточным базисом

### 4.3. Комбинированная система кодирования широкополосной речевой информации

В настоящем разделе представлена комбинированная система кодирования с встроенным препроцессором повышения качества речи (см. рис. 4.9). Гибридная система представляет собой широкополосный кодер, описанный в п.4.1, и систему подавления шума, представленную в п.4.2. Система препроцессора была немного модифицирована для соответствия требованиям кодера, предъявляемым к сегментации входного сигнала. Таким образом, система подавления шума разбивает входной сигнал на фреймы  $N=320$  отсчетов (20мс) с 50% перекрытием и взвешивает входной сигнал на окно Хэмминга. Для обеспечения совместимости, был выбран размер окна анализа WDFT длиной в 320 отсчетов. В соответствии с формулой (3.25), для  $N=320$  и  $\alpha = -0.57827$ , размер избыточного базиса WDFT составит  $M_{opt} \approx 1197$ . Учитывая вычислительную сложность, величина избыточного базиса была снижена до  $M=512$ .

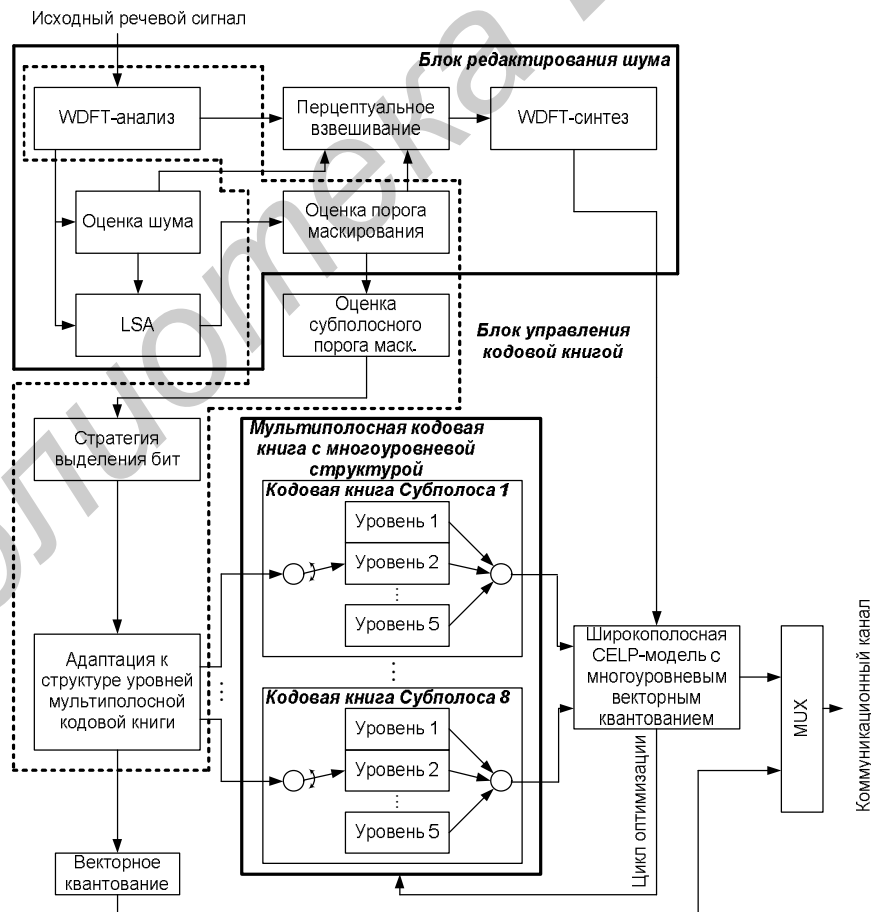


Рис. 4.9. Комбинированная система кодирования

В блоке WDFТ синтеза осуществляется формирование обработанного фрейма сигнала путем перекрытия с накоплением. Очередной сформированный фрейм поступает на вход кодера (см. рис. 4.10). Как видно из рис. 4.10, блок редактирования шума добавляет алгоритмическую задержку равную половине фрейма (10мс), что приводит к общей алгоритмической задержке комбинированной системы кодирования равной 30мс.

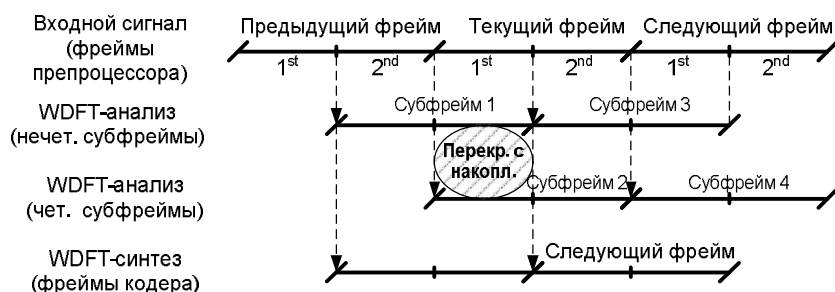


Рис. 4.10. Организация вычислительного процесса WDFТ анализа-синтеза

В рассматриваемой системе принята следующая конфигурация. Блок системы шумоподавления использует WDFТ с базисом 320x512. Для правила  $H^{IND}$  остаточный уровень шума зафиксирован на уровне  $\zeta_d = 0.05$ . Правило LSA сконфигурировано для достижения максимального шумоподавления в 20дБ и работы с шумами выше -10дБ. Блок мониторинга структуры кодовой книги реконфигурирует структуру кодера каждый нечетный субфрейм WDFТ анализа (см. рис. 4.10). Схема кодирования параметров кодера аналогична представленной в табл. 4.2.

В качестве примера на рис. 4.11 представлен сравнительный анализ спектрограмм оригинального речевого фрагмента, зашумленного, очищенного (системой подавления шума) и восстановленного после кодирования комбинированной системой.

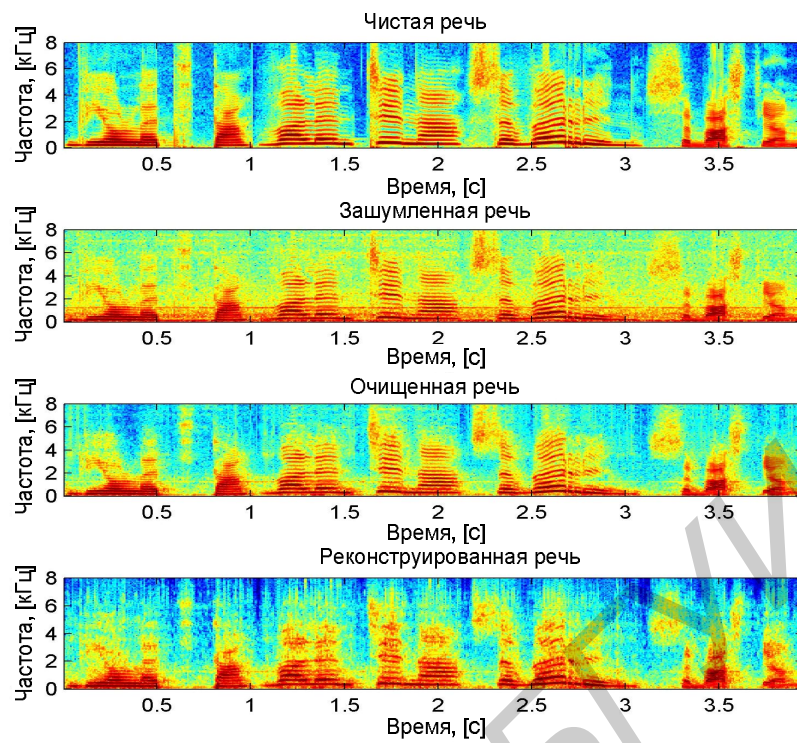


Рис. 4.11. Сравнение спектрограмм для аддитивного шума типа «шум автомобиля» с уровнем 5дБ

## Литература

### *Матрицы, матричный анализ*

1. Хорн Р., Джонсон Ч. Матричный анализ. – М.: Мир, 1989. – 655с.
2. Форсайт Дж., Малькольм М., Моулер К. Машинные методы математических вычислений. – М.: Мир, 1980. – 280с.
3. Penrose R. A Generalized Inverse for Matrixes, *Proc. Cambridge Phil. Soc.* 51, 1955, pp. 406-413.

### *Цифровая обработка речевых сигналов*

4. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов. – М.: Радио и связь, 1981. – 495 с.
5. Назаров М.В., Прохоров Ю.Н. Методы цифровой обработки и передачи речевых сигналов. –М.: Радио и связь, 1985.
6. Прохоров Ю.Н. Статистические модели и рекуррентное предсказание речевых сигналов. –М.: Радио и связь, 1984.
7. Применение цифровой обработки сигналов/ под редакцией Э. Оппенгейма. –М.: Мир, 1980.

### *Цифровые банки фильтров*

8. Crochiere R.E., Rabiner L. Multirate Digital Signal Processing. Prentice Hall, Englewood Cliffs, NJ, 1983.
9. Vaidyanathan P.P. Multirate systems and filter banks. Prentice Hall, Englewood Cliffs, 1993.
10. Straug G., Nguyen T. Wavelets and filter banks. –Wellesley: Wellesley-Cambridge Press, 1996.
11. Витязев В.В. Цифровая частотная селекция сигналов. –М.:Радио и связь, 1993, 240с.
12. Крот А.М., Кудрявцев В.О. Теория анализа и синтеза бэнк-фильтров и их применение // Зарубежная радиоэлектроника, 1999, №2, с.3-17.
13. Зайцев А.А. Методы построения банков цифровых фильтров: тематический обзор // Цифровая обработка сигналов, 2003, №1, с.2-10.
14. Гусинская Е.И., Зайцев А.А. Оптимизация банка фильтров в задачах субполосного кодирования: тематический обзор // Цифровая обработка сигналов, 2004, №3, с.18-28.
15. Вайдьянатхан П.П. Цифровые фильтры, блоки фильтров и полифазные цепи с многочастотной дискретизацией: Методический обзор // ТИИЭР, 1990, №3, с.77-119.
16. Теория и применение цифровой обработки сигналов: Лаб. практикум по синтезу цифровых фильтров с конечной импульсной характеристикой для студ. спец. 40.02.02 «Электронные вычислительные средства» дневной формы обучения / А.А. Петровский, Д.С. Лихачёв, Ал.А. Петровский. – Мн.: БГУИР, 2004. – 55 с.: ил.32.



17. Kok C.W., Nguyen T.Q. Multirate filter banks and transform coding gain. *IEEE Transactions Signal Processing*, 46(7). July 1998, pp. 2041-2044.
18. Gosse K., Duhamel P. Perfect reconstruction versus MMSE filter banks in source coding. *IEEE Transactions Signal Processing*, 45(9), September 1997, pp. 2188-2202.
19. Smith J.O. and Abel J.S. Bark and ERB Bilinear Transforms. *IEEE Trans. Speech and Audio Proc.*, 7(6), 1999, pp. 697-708.
20. Petrovsky A., Parfieniuk M., Bielawski K. Psychoacoustically Motivated Non-uniform Cosine Modulated Polyphase Filter Bank. *2nd International Workshop on Spectral Methods and Multirate Signal Processing (SMMSP 2002)*, 2002 pp.95-101, Toulouse, France, September 7-8.
21. Sathe V.P., Vaidyanathan P.P. Effects of multirate systems on the statistical properties of random signals. *IEEE Transactions Signal Processing*, 41(1). January 1993, pp. 131-146.
22. Soman A.K., Vaidyanathan P.P. Coding gain in paraunitary analysis/synthesis systems. *IEEE Transactions Signal Processing*, 41(5), May 1993, pp. 1824-1835.
23. Djokovic I., Vaidyanathan P.P. On optimal analysis/synthesis filters for coding gain maximization. *IEEE Transactions Signal Processing*, 44(5). May 1996, pp. 1276-1279.
24. Calvagno G., Mian G.A., Rinaldo R. Computation of the coding gain for subband coders. *IEEE Transactions Communications*, 44(4). April 1996, pp. 475-487.
25. Uzun N., Haddad R.A. Cyclostationary modeling, analysis, and optimal compensation of quantization errors in subband codecs. *IEEE Transactions Signal Processing*, 43(9), September 1995, pp. 2109-2119.
26. Papoulis A. Probability, random variables, and stochastic processes. –McGraw-Hill, Tokyo, 1984.

#### **Компрессия речевых сигналов**

27. Сапожков М.А., Михайлов В.Г. Вокодерная связь. –М.: Радио и связь, 1983.
28. Шелухин О.И., Лукьянцев Н.Ф. Цифровая обработка и передача речи. –М.: Радио и связь, 2000.
29. Речевые интерфейсы ЭВС: Учебно-метод. пособие для студ. спец. I – 40 02 02 «Электронные вычислительные средства» дневной формы обуч. / А.А. Петровский, Д.С. Лихачёв, Ал.А. Петровский и др. – Мн.: БГУИР, 2004. – 51 с.: ил.
30. Речевой интерфейс ЭВС: Метод. пособие для студ. спец. I – 40 02 02 «Электронные вычислительные средства» дневной формы обучения / А.А. Петровский, Ал.А. Петровский, Д.С. Лихачев. – Мн.: БГУИР, 2004. – 55 с.: ил.
31. Kondo A.M. Digital Speech Coding for Low Bit Rate Communications Systems, UK, University of Surrey: John Wiley & Sons, 1996 – 442 p.

32. Speech coding and synthesis// Edited by W.B.Kleijn, K.K.Paliwal. Amsterdam, Elsevier, 1998.
33. Barnwell T.P., Nayebi K., Richardson C.H. Speech coding: a computer laboratory textbook, NJ, J.Wiley & Sons, 1996.
34. Hess W.J. Pitch and voicing determination. // Advances in speech signal processing, Sadaoka Furui and M. M. Sohndi, Eds. New York: Marcel Dekker, 1992, pp. 3-48.
35. Mobile Radio Communications /by Raymond Steele edition, PENTECH PRESS Publishers – London, 1992.
36. Zwicker E., Fastl H. Psychoacoustics: Facts and Models. Springer-Verlag Berlin. Heidelberg, 1990.
37. Johnston J.D. Transform coding of audio signals using perceptual noise criteria // IEEE Trans. on Select. Areas Commun. Feb. 1988, vol. 6, pp. 314-323.
38. Jayant N.S., Noll P. Digital coding of waveforms. Prentice Hall: Englewood Cliffs, 1984.

#### **Редактирование шума в речевом сигнале**

39. Makur A., Mitra S.K. Warped Discrete-Fourier Transform: Theory and Applications. *IEEE Trans. Circuits and Systems I*, 2001, vol. 48, №9, pp. 1086-1093.
40. Lim J.S. Speech Enhancement. Prentice-Hall, 1983.
41. Vaseghi S.V. Advanced Digital Signal Processing and Noise Reduction. Second Ed., Wiley, 2000.
42. Collura J.S. Speech Enhancement and Coding in Harsh Acoustic Noise Environments. *IEEE Workshop on Speech Coding*. Haikko Manor, Porvoo, Finland, 1999.
43. Gustafsson S., Martin R., Jax P., Vary P. A psychoacoustic approach to combined acoustic echo cancellation and noise reduction. *IEEE Trans. on Speech and Audio Processing*, 2002, vol. 10, №5, pp. 245-256.
44. Parfieniuk M., Petrovsky A. Warped DFT as the basis for psychoacoustic model. *Proc. of the IEEE International conference on Acoustic, Speech, Signal processing, ICASSP*, vol. IV, Montreal, Canada, May 2004, pp. 185-188.
45. Petrovsky A., Parfieniuk M., Borowicz A. Warped DFT based perceptual noise reduction system. *Proc. AES 116th*, Berlin, Germany, 2004, Conv. Paper #6035.
46. Ephraim Y., Malah D. Speech enhancement using a minimum mean-square log-spectral amplitude estimator. *IEEE Trans. on Acoustic Speech and Signal Processing*, vol. ASSP-33, 1985, №2, pp. 443-445.
47. Cohen I., Berdugo B. Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement. *IEEE Signal Processing Letters*, 2002, vol. 9, №1, pp. 12-15.
48. Borowicz A., Petrovsky A. An Overcomplete WDFT Sinusoidal Basis for Perceptually Motivated Speech Enhancement. *Proc. of 13th European Signal Processing Conference (EUSIPCO'2005)*, (CD-ROM), Antalya, 2005.

*Учебное издание*

**Петровский Александр Александрович,  
Парфенюк Марек,  
Борович Адам,  
Лившиц Михаил Зенадьевич**

# **ЦИФРОВЫЕ БАНКИ ФИЛЬТРОВ: АНАЛИЗ, СИНТЕЗ И ПРИМЕНЕНИЕ В МУЛЬТИМЕДИА СИСТЕМАХ**

## **УЧЕБНО-МЕТОДИЧЕСКОЕ ПОСОБИЕ**

по курсам  
«Теория и применение цифровой обработки сигналов»  
и «Речевые интерфейсы ЭВС»  
для студентов специальности I – 40 02 02  
«Электронные вычислительные средства»  
дневной формы обучения

Ответственный за выпуск М.З. Лившиц

Подписано в печать	Формат 60x84 1/16.	Бумага офсетная.
Гарнитура «Таймс»	Печать ризографическая.	Усл. печ. л.
Уч.- изд. л. 4,0	Тираж 200 экз.	Заказ 365

Издатель и полиграфическое исполнение: Учреждение образования  
«Белорусский государственный университет информатики и радиоэлектроники»  
ЛИ №02330/0056964 от 01.04. 2004. ЛП №02330/0131518 от 30.04. 2004.  
220013, Минск, П. Бровки, 6.