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

Кафедра электронных вычислительных средств

А.А.Петровский, Ал.А.Петровский, Д.С.Лихачев

РЕЧЕВЫЕ ИНТЕРФЕЙСЫ ЭВС

МЕТОДИЧЕСКОЕ ПОСОБИЕ

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

Минск 2004

УДК 004.5 (075.8)
ББК 32.973 я 73
П 30

Речевые интерфейсы ЭВС: Метод. пособие для студ. спец. 40 02 02 П 30 “Электронные вычислительные средства” дневной формы обуч. / А.А. Петровский, Ал.А. Петровский, Д.С. Лихачев. – Мн.: БГУИР, 2004. – 55 с.: ил.
ISBN 985-444-643-3

В данном методическом пособии рассматриваются методы компрессии речевых сигналов с психоакустической мотивацией на основе линейного предсказания и пакета дискретного вэйвлет-преобразования (ПДВП). Показаны алгоритмическое обеспечение и программная модель широкополосных вокодеров на основе CELP-модели с многополосным возбуждением и перцептуальной оптимизацией. Описан перцептуальный широкополосный кодер речевых сигналов на основе ПДВП с процедурой расчета психоакустической модели в вэйвлет-области.

Авторы выражают благодарность аспиранту кафедры ЭВС М.З. Лившицу за помощь в подготовке и проведении модельных экспериментов с широкополосным CELP-кодером.

УДК 004.5 (075.8)
ББК 32.973 я 73

ISBN 985-444-643-3

ВВЕДЕНИЕ

Непрерывное увеличение передач аудио- и речевых данных в системах мультимедиа через Интернет обусловливает поиск новых решений цифровой обработки в реальном масштабе времени аудио- и речевых сигналов (их компрессию и декомпрессию). Использование цифрового представления данных позволяет обеспечить надежность и экономичность связи, возможность гарантированной защиты от несанкционированного доступа. Благодаря появлению новых процессоров обработки сигналов появилась возможность быстрого создания перспективных систем передачи или запоминания речевых сообщений. При этом большое значение приобретает проблема минимизации числа бит, необходимых для передачи сигнала, проблема кодирования и компрессии речи. Актуальной задачей обработки речи стало создание систем низкоскоростной передачи с высоким качеством восприятия сигнала, способных функционировать в реальных условиях. В системах речевой связи данные передаются, хранятся и обрабатываются различными способами, что обусловливает применение разных форм представления речевого сигнала. К ним предъявляются следующие требования:

- сохранение информационного содержания речи;
- удобство передачи и хранения;
- возможность легкого и гибкого преобразования без существенных информационных потерь (задача кодирования и декодирования речевого сигнала).

Целью цифрового кодирования речи является получение как можно более высокого качества восстанавливаемой речи при наименьшей скорости передачи. Исследования в области кодирования речи проводятся более трех десятков лет. Результат этих интенсивных усилий – множество различных стратегий и подходов для кодирования речевых сигналов, ряд которых доведен до соответствующих национальных и международных стандартов. Классификация различных стратегий и схем цифрового кодирования речи показана на рис. В1, а краткий обзор способов их построения приведен в прил. 3 и 4.

Схемы кодирования речи характеризуются следующими параметрами:

- скоростью передачи;
- вычислительной и емкостной сложностью технической реализации;
- сложностью алгоритма;
- задержкой сигнала;
- качеством восстановленной речи. Данный параметр является функцией от первых трех, и, как правило, повышение качества предполагает увеличение первых трех параметров.

Для большинства приложений некоторые из характеристик предопределены. Например, коммуникационный канал может накладывать ограничения на скорость передачи данных. Качество обычно может быть улучшено при увеличении скорости, сложности, а иногда и при увеличении задержки. В телефонной сети речь фильтруется полосовым фильтром с полосой пропускания 200...3200 Гц, это так называемая трёх килогерцовая речь, которая оцифровывается с частотой

дискретизации 8 кГц. В настоящее время наибольший интерес представляет передача широкополосной речи (50...7000 Гц), частота дискретизации которой удвоена до 16 кГц.

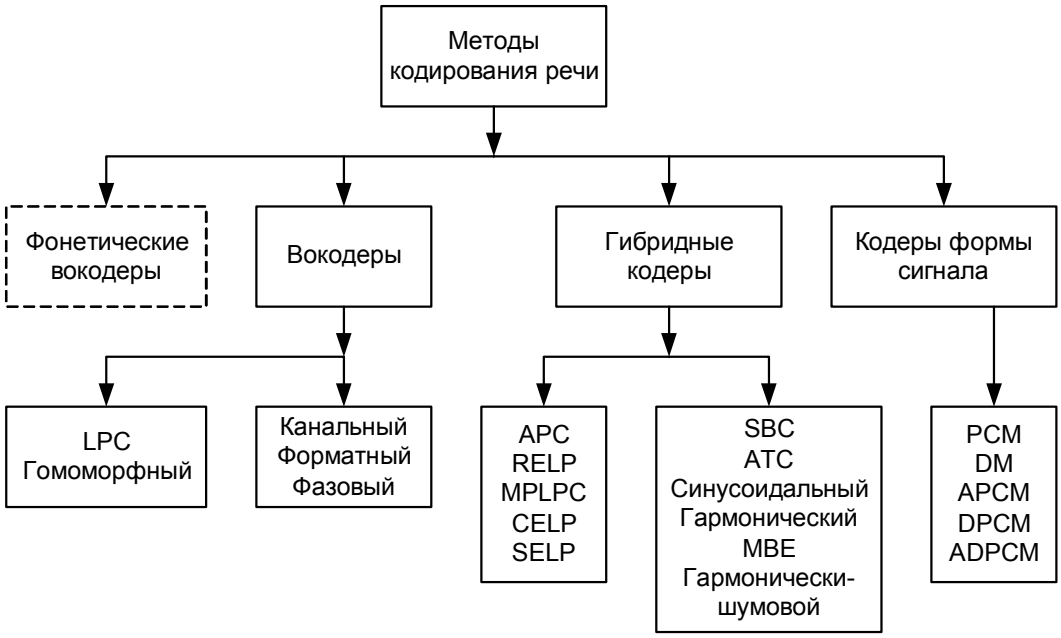


Рис. В1. Классификация методов кодирования речи

Скорость передачи. Количество бит, передаваемых в единицу времени, характеризует степень компрессии, которая достигается кодером. В телефонии речь оцифровывается с частотой 8 кГц и квантуется 8-битным логарифмическим квантователем, таким образом, результирующая скорость составляет 64 Кбит/с. Для кодеров телефонной полосы частот степень компрессии определяет, насколько скорость потока данных меньше, чем 64 Кбит/с. Речевые кодеры должны иметь постоянную скорость. Степень компрессии может быть увеличена, если исключить передачу данных во время паузы или молчания в диалоге абонентов.

Задержка сигнала. Коммуникационная задержка кодера более важна для передачи данных, нежели для их хранения. Большая коммуникационная задержка может приводить к неудобству или даже к невозможности нормального общения абонентов. Большинство низкоскоростных кодеров являются блочными. Они кодируют блок речи (фрейм) в единицу времени. Задержка при кодировании речи определяется, во-первых, алгоритмической задержкой – временем получения фрейма сигнала, во-вторых, вычислительной задержкой – временем обработки фрейма, которое зависит от производительности используемого процессора. Другие задержки в законченных системах – это задержка мультиплексирования и передачи данных в линию.

Сложность технической реализации. Степень сложности определяют два основных фактора – цена и потребляемая мощность. Цена, как правило, зависит от выбранного типа процессора для данного приложения. 16-битные цифровые процессоры обработки сигналов (ЦПОС) или digital signal processing (DSP-процессоры) в основном предпочтительнее для реализации речевых кодеров, так как последние менее дорогие и потребляют меньше мощности, чем аналогичные

реализации на DSP-процессорах с плавающей запятой. Недостатком ЦПОС с фиксированной запятой является то, что алгоритмы кодирования речи должны быть реализованы с использованием целочисленной 16-битной арифметики. В то же время кодеры речи на DSP-процессорах с плавающей запятой характеризуются высокой арифметической точностью, приблизительно такой же, как и у симуляторов на высокоуровневых языках. Следовательно, эффект конечной разрядности регистров ЦПОС здесь может не учитываться.

Качество восстановленной речи. Атрибут качества имеет много параметров. В конечном счете качество определяется тем, как реконструированная декодером речь воспринимается слушателем. Чистая или зашумленная речь, ошибки в потоке данных являются одними из многих факторов, влияющих на качество кодирования. Оценка качества компрессоров речевых сигналов определяется по результатам субъективных испытаний. Эксперты прослушивают пары предложений и дают одну из следующих оценок: отлично, хорошо, удовлетворительно, плохо. Типичный тест содержит множество фраз, произнесенных большим количеством дикторов.

1. ОСОБЕННОСТИ ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ

1.1. Характеристика речевых сигналов

Речевое сообщение начинается с того, что в мозгу диктора возникает в абстрактной форме некоторая фраза. В процессе речеобразования она преобразуется в акустическое колебание. Сообщение, передаваемое с помощью речевого сигнала, является дискретным, т.е. может быть представлено в виде последовательности фонем из конечного их числа. Для иллюстрации особенностей речевого сигнала фраза «Кафедра ЭВС», произнесенная мужчиной, была оцифрована, введена в компьютер и обработана в среде MATLAB. На рис. 1.1 приведены временное (вверху) и спектральные представления фразы с меньшим (в середине) и большим (внизу) разрешением по частоте. Временной масштаб всех графиков согласован.

Спектрограмма описывает энергию сигнала в координатах «время – частота – яркость», затемненные участки соответствуют областям концентрации энергии. При малом спектральном разрешении из-за сглаживания точность оценки частотных параметров сигнала мала. Временные же характеристики (например, границы слов и отдельных фонем) могут быть определены достаточно точно. Большее разрешение в частотной области достижимо только за счет ухудшения разрешения по времени. На нижнем рисунке узкие горизонтальные линии на спектрограмме соответствуют траекториям гармоник основной частоты. Для невокализованных звуков подобной структуры спектра не наблюдается.

Особый интерес представляет оценка скорости передачи информации, содержащейся в речевом сигнале. Нижний предел определяется скоростью произнесения фонем (в среднем 10 фонем в секунду) и составляет 50-60 бит/с. Эта оценка не учитывает таких факторов, как индивидуальность и эмоциональное состояние диктора, громкость речи и т.д. Общая же скорость передачи речевого сигнала (но не его информационного содержания) при определенном представлении может лежать в диапазоне от 400 до 10^6 бит/с.

Основная сложность при разработке метода компрессии речи заключается, во-первых, в выборе модели речеобразования. Модель должна учитывать характеристики речевого сигнала, оказывающие наибольшее влияние на качество восстановления. Во-вторых, достаточно сложно определить способ квантования параметров модели, который и обусловливает требуемую скорость передачи. Длительное время в алгоритмах кодирования эти вопросы решались с точки зрения минимизации расхождения между исходным и синтезированным речевым сигналом во временной либо в частотной области и не учитывались особенности восприятия речи человеком. Между тем качество самих алгоритмов всегда оценивается субъективно. Работы последних лет по психоакустике показали, что использование психоакустических закономерностей дает возможность достижения большей степени компрессии речевых сигналов. При этом выбираются такие параметры речевого сигнала и таким образом квантуются, чтобы обеспечить искажения, минимальные для слушателя.

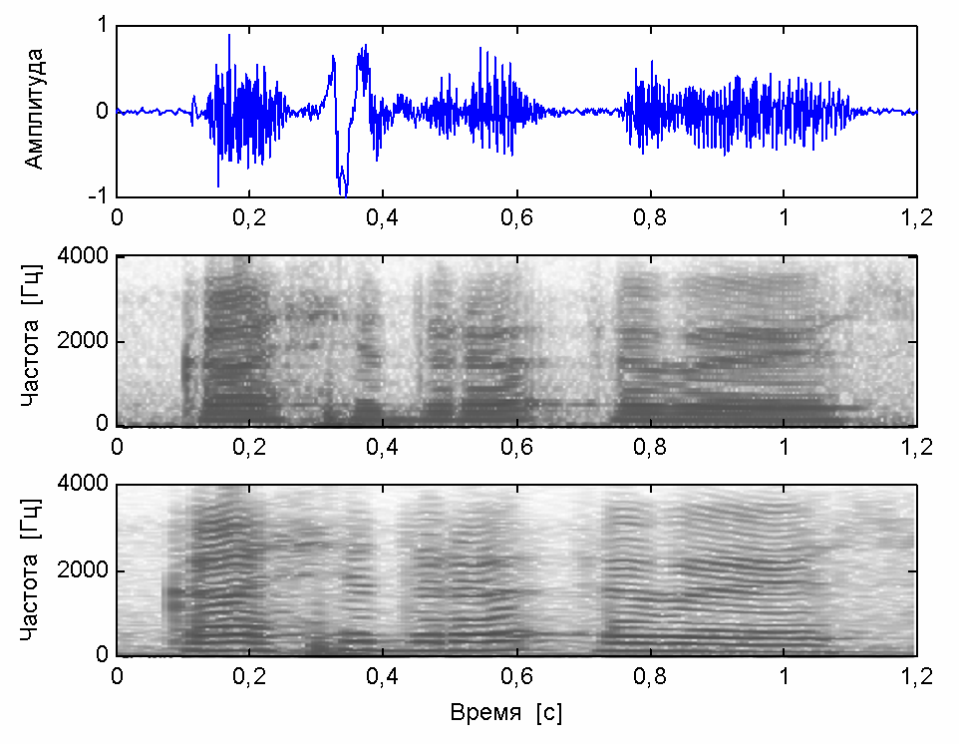


Рис. 1.1. Пример речевого сигнала

1.2. Принципы психоакустики

1.2.1. Абсолютный порог слышимости

Исследования в области психоакустики ведутся на протяжении нескольких последних десятилетий. Наиболее полное изложение накопленных фактов и разработанных моделей содержится в работе: Zwicker E., Fastl H. Psychoacoustics: Facts and Models. Springer-Verlag Berlin Heidelberg, 1990. Основными результатами исследований, которые нашли применение в цифровой обработке речевых сигналов, являются абсолютный порог слышимости, критические полосы, частотное и временное маскирование.

Под абсолютным порогом слышимости понимают минимальную энергию чистого тона, которая еще позволяет слушателю определить наличие этого тона при отсутствии окружающего шума. Зависимость этой энергии от частоты, показанная на рис. 1.2. Для аппроксимации частотной зависимости абсолютного порога слышимости хорошо подходит следующая нелинейная функция:

$$T(f) = 3{,}64(f / 1000)^{-0{,}8} - 6{,}5 \exp(-0{,}6\,(f / 1000 - 3{,}3)^2) + 10^{-3}(f / 1000)^4 \text{ [дБ]}, \qquad (1.1)$$

где f – частота в [Гц].

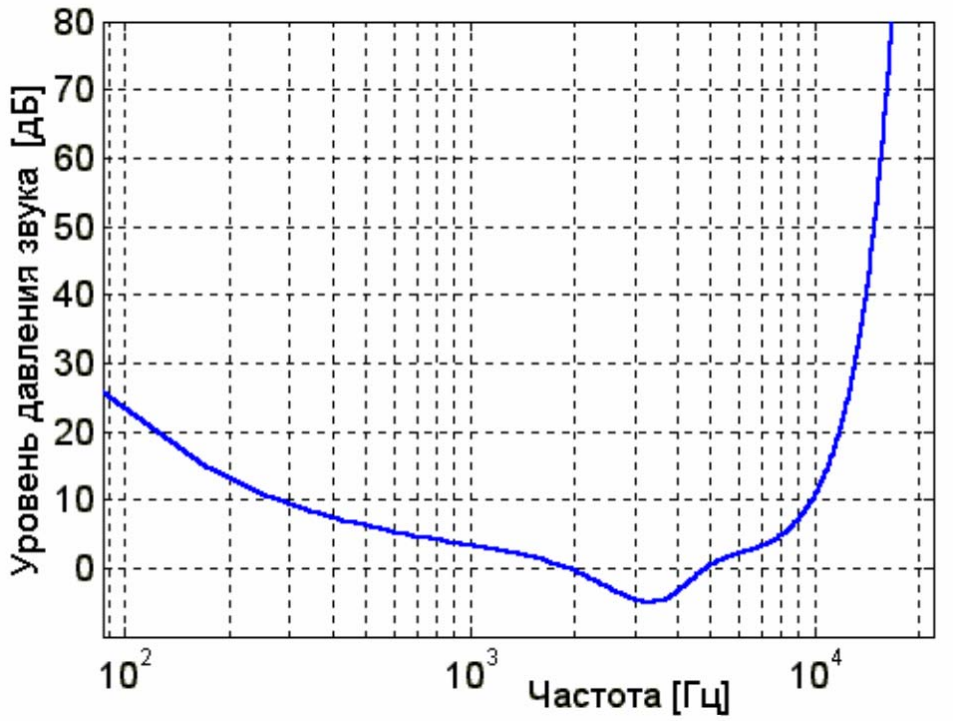


Рис. 1.2. Абсолютный порог слышимости

В задачах компрессии речи под $T(f)$ можно понимать максимально возможный уровень искажений гармонического сигнала данной частоты, не сказывающийся на качестве. Обычно абсолютный порог слышимости используется для взвешивания речевого сигнала в частотной области с целью выравнивания воспринимаемого слушателем уровня погрешностей кодирования по частоте.

1.2.2. Критические полосы восприятия акустической информации

Помимо чувствительности слуха от частоты зависит и его спектральное разрешение. В физиологии это объясняется преобразованием частота – место во внутреннем ухе. Различные участки в улитке, каждый со своими рецепторами, “настроены” на разные частотные полосы, которые называют критическими. Эмпирические работы ряда исследователей подтвердили соответствие критических полос отдельным участкам улитки. Экспериментально ширина критической полосы может быть определена по резкому уменьшению субъективной громкости. На рис. 1.3,а схематически показан пример такого определения. Узкополосный шумовой сигнал маскировался двумя близкорасположенными тональными компонентами, расстояние между которыми постепенно увеличивалось. На

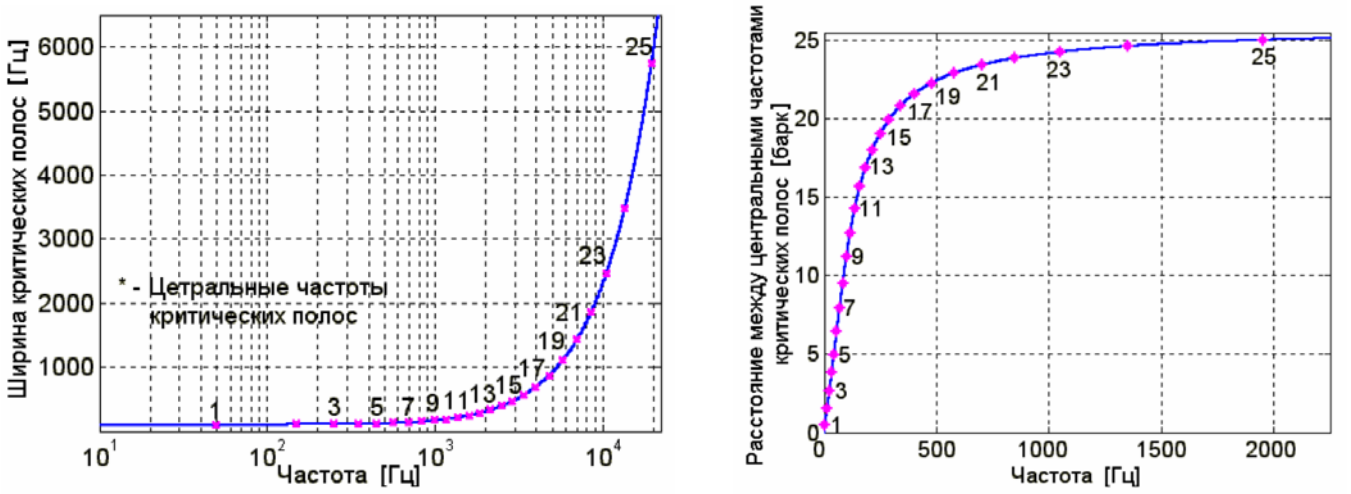
рис. 1.3,б приведено поведение порога обнаружения сигнала. В пределах ширины критической полосы $f_{кп}$ он остается постоянным, а за ее пределами резко уменьшается.

а) б)

Рис. 1.3. Схема определения ширины критических полос:
а – узкополосный шумовой компонент маскируется двумя тоновыми компонентами;
б – зависимость поведения порога обнаружения сигнала от изменения расстояния между двумя тоновыми компонентами Δf.

Ширина критической полосы остается примерно постоянной (около 100 Гц) вплоть до значения центральной частоты полосы 500 Гц, а при больших значениях увеличивается в среднем на 20% центральной частоты, как это показано на рис. 1.4,а. В работах по психоакустике используется следующая аппроксимация этой зависимости:

$$BW(f) = 25 + 75\ (\ 1 + 1{,}4\ (f / 1000)^2)^{0{,}69}\ [\text{Гц}]. \qquad (1.2)$$



а) б)

Рис. 1.4. Частотные зависимости:
а – ширины критических полос; б – центральных частот критических полос

Хотя функция *BW* непрерывна, для практических задач лучше использовать дискретный набор полосовых фильтров, покрывающий всю рабочую полосу частот (табл. 1.1). Для определения расстояния между центральными частотами соседних критических полос часто используют специальную единицу частоты: 1 барк. На

рис. 1.4,б показана связь между частотой, выраженной в герцах и барках. Эта зависимость используется для преобразования линейной шкалы частот в шкалу, соответствующую спектральному восприятию человека. Для ее аппроксимации было получено следующее выражение:

$$z(f) = 13 \arctan(0{,}00076\, f) + 3{,}5 \arctan((f/7500)^2) \text{ [барк]}. \quad (1.3)$$

Таблица 1.1.

№ полосы	Центральная частота [Гц]	Ширина полосы [Гц]	№ полосы	Центральная частота [Гц]	Ширина полосы [Гц]
1	50	0…100	14	2150	2000…2320
2	150	100…200	15	2500	2320…2700
3	250	200…300	16	2900	2700…3150
4	350	300…400	17	3400	3150…3700
5	450	400…510	18	4000	3700…4400
6	570	510…630	19	4800	4400…5300
7	700	630…770	20	5800	5300…6400
8	840	770…920	21	7000	6400…7700
9	1000	920…1080	22	8500	7700…9500
10	1170	1080…1270	23	10500	9500…12000
11	1370	1270…1480	24	13500	12000…15500
12	1600	1480…1720	24	19500	15500…22050
13	1850	1720…2000			

1.2.3. Маскирование

С механизмом критических полос человеческого слуха связаны свойства межполосового и внутриполосового частотного маскирования. Под маскированием понимают ситуацию, при которой один звук становится неслышимым из-за присутствия другого звука. С целью оптимального распределения погрешностей кодирования следует различать два вида частотного маскирования: тон–шум и шум–тон. В первом случае тоновой сигнал, расположенный в центре критической полосы, маскирует шум в пределах ширины полосы или некоторой ее окрестности; спектр шума оказывается ниже порога обнаружения, определяемого маскирующим тоном. Во втором случае, наоборот, маскирующим сигналом является шум, а маскируемым–тон.

Эффект маскирования упрощенно можно объяснить тем, что сильный тоновой или шумовой маскер создает очаг возбуждения на участке базилярной мембраны, соответствующем критической полосе. Это возбуждение препятствует передаче более слабого сигнала. Порог маскирования снижается при увеличении разницы частот маскирующего и маскируемого сигналов. Данное явление, называемое распространением маскирования, в алгоритмах кодирования часто моделируется треугольной функцией распространения с наклоном +25 и –10 дБ/барк. На рис. 1.5,а показан расположенный в середине критической полосы маскирующий тоновый сигнал и соответствующий ему порог маскирования. Для более точного представления функции распространения используется выражение

$$SF(x) = 15{,}81 + 7{,}5(x + 0{,}474) - 17{,}5\sqrt{1 + (x + 0{,}474)^2} \ \text{[дБ]}, \qquad (1.4)$$

где x – разность частот маскирующего и маскируемого сигналов в [барк].

а) б)

Рис. 1.5. Маскирование: а – частотное; б – временное;

Величина порогов маскирования может быть определена следующим образом:

$$TH_ш = E_m - 14{,}5 - B, \qquad TH_m = E_ш - K, \qquad (1.5)$$

где $TH_ш$ и TH_m – пороги маскирования тон–шум и шум–тон в [дБ]; E_m и $E_ш$ – уровни энергии тонового и шумового маскирующих сигналов в критической полосе; B – номер критической полосы.

В зависимости от алгоритма параметр K принимает значение от 3 до 5 дБ. Маскирование имеет место и во временной области. Наличие громкого сигнала с резкими временными границами создает области пре- и пост-маскирования, в течение которых слушатель не воспринимает другие сигналы с энергией, не превышающей порога маскирования. Это явление схематически показано на рис. 1.5,б. Интервал пре-маскирования составляет примерно 5 мс, а пост-маскирования – от 50 до 300 мс в зависимости от силы и длительности маскера. Временное маскирование может с успехом использоваться в вокодерах с переменной скоростью передачи.

1.3. Методы обработки речевых сигналов на основе принципов психоакустики

Один из первых методов, анализ речевого сигнала, которого основывался на последовательном применении принципов психоакустики, был метод PLP (Perceptual Linear Prediction). Классическая схема линейного предсказания была преобразована с учетом особенностей восприятия речи человеком. Для построения модели используется «слышимый спектр». Структурная схема метода приведена на рис. 1.6. С помощью кратковременного преобразования Фурье вычисляется спектр мощности входного сигнала. Частотная шкала спектра модифицируется преобразованием герц–барк. Для выделения критических полос использует набор взвешивающих функций, представленный на рис. 1.7. Данные взвешивающие функции были получены по оригинальной аппроксимации асимметричных маскирующих кривых с учетом кривой равной громкости. Полученный спектр в

критических полосах позволяет снизить частотное разрешение до 18 спектральных отсчетов с интервалом около 1 барк. Классические алгоритмы анализа речи строят модели по логарифму спектра мощности. В отличие от этой практики было предложено использовать преобразование интенсивность–громкость, примерно соответствующее закону кубического корня. Описанная процедура анализа применялась для задач распознавания речи, обеспечивая в некоторых случаях снижение количества ошибок почти на 20% по сравнению с традиционным линейным предсказанием.

Рис. 1.6. Структурная схема PLP

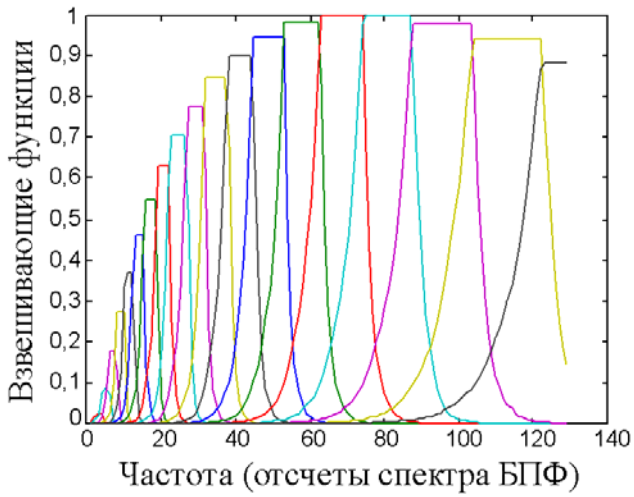


Рис. 1.7. Взвешивающие функции PLP

Большой интерес представляет также метод инструментального измерения качества синтезированной речи. При этом ставится задача получить ту же оценку, которую бы выставил при прослушивании человек. Базовая структура алгоритма представлена на рис. 1.8. Погрешность между исходным $x(k)$ и синтезированным $y(k)$ сигналом определяется по результатам психоакустического анализа (рис. 1.9). Реализация данного метода позволила достичь высокой корреляции ($\rho \approx 0{,}95$) между субъективными и инструментальными результатами.

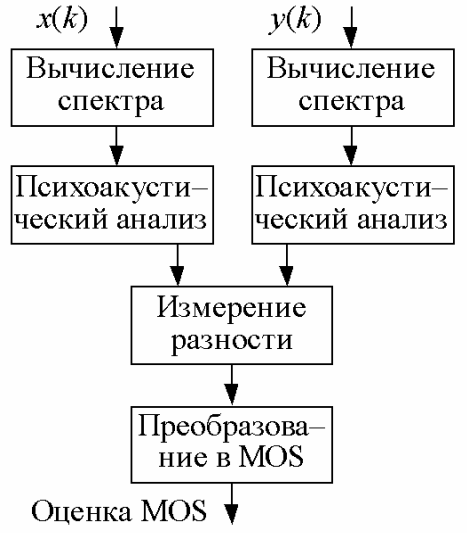


Рис. 1.8. Структура метода инструментального измерения качества речи

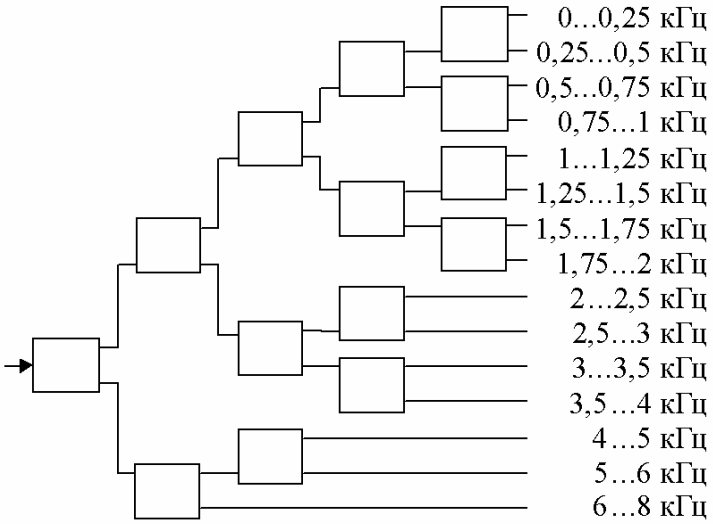


Рис. 1.9. Психоакустический банк фильтров

В последние годы элементы психоакустики находят широкое применение для компрессии аудио-сигналов, когда на основе зависимостей ширины и центральной частоты критических полос адаптируется банк анализирующих фильтров. В комбинированных же системах подавления эха сигнала и шума используется порог восприятия акустической информации человеком для взвешивания речевого сигнала в частотной области. При этом коэффициент возврата эха в канал составляет более минус 40 дБ.

2. КОМПРЕССИЯ РЕЧЕВЫХ СИГНАЛОВ С ПСИХОАКУСТИЧЕСКОЙ МОТИВАЦИЕЙ НА ОСНОВЕ СХЕМЫ "АНАЛИЗ ЧЕРЕЗ СИНТЕЗ"

2.1. Модель кодирования на основе линейного предсказания по схеме "анализ через синтез"

2.1.1. Общая структура модели

Общая структура модели кодирования речи на основе линейного предсказания, выполненная по схеме "анализ через синтез", представлена на рис. 2.1.

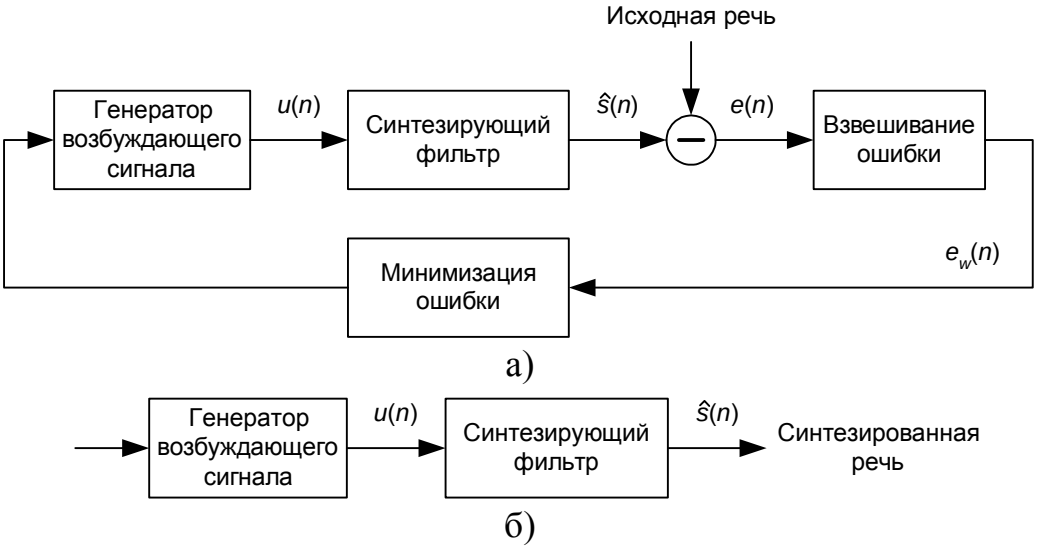


Рис. 2.1. Общая модель кодирования на основе линейного предсказания по схеме "анализ через синтез": а – кодер; б – декодер

Модель состоит из трех основных частей.

Первая часть – это синтезирующий фильтр, который является полюсным подстраиваемым фильтром для моделирования кратковременной спектральной огибающей речевого сигнала. Его часто называют кратковременным корреляционным фильтром (*"short-term correlation filter (predictor)" - STP*), так как коэффициенты рассчитываются предсказанием очередного отсчета по нескольким предыдущим. Синтезирующий фильтр может также включать долговременной корреляционный фильтр (*"long-term correlation filter (predictor)" - LTP*), включенный последовательно с кратковременным корреляционным фильтром.

Вторая часть модели – генератор возбуждения, который выдает возбуждающую последовательность на вход синтезирующего фильтра для получения реконструированной речи на стороне приемника. Сигнал возбуждения оптимизируется посредством минимизации перцептуально взвешенной ошибки между оригинальной и синтезированной речью. Эффективность этого метода обусловлена оптимизационной процедурой замкнутого цикла, которая позволяет представлять предсказанный сигнал, малым количеством бит при сохранении высокого качества реконструированной речи.

Третья часть модели – модуль минимизации ошибки. В этой части модели используется субъективно значимый критерий, где ошибка $e(n)$ пропускается через перцептуальный взвешивающий фильтр, который окрашивает шумовой спектр и концентрирует энергию на формантных частотах речевого спектра, т.е. шум маскируется речевым сигналом.

Процедура кодирования включает два шага. Сначала определяются параметры синтезирующего фильтра по отсчетам речевого сигнала (10…30 мс) вне цикла оптимизации, затем определяется оптимальная возбуждающая последовательность, минимизирующая критерий взвешенной ошибки. Интервал (длительность) оптимизируемого сигнала возбуждения обычно составляет 4…7,5 мс, что меньше, чем интервал обновления параметров LPC. Поэтому речевой фрейм делится на субфреймы (субблоки), на которых сигнал возбуждения определяется индивидуально. Квантованные параметры фильтра и сигнала возбуждения пересылаются на сторону приемника.

Процедура декодирования осуществляется путем пропускания декодированного сигнала возбуждения через синтезирующий фильтр для получения на выходе реконструированной речи.

2.1.2. Кратковременной фильтр-предсказатель STP

Роль STP-фильтра заключается в представлении огибающей спектра речи. В CELP-синтезаторе идеально плоский сигнал возбуждения окрашивается огибающей STP-фильтра. Параметры STP могут определяться различными методами. Наиболее удобным и получившим распространение является метод на основе автокорреляции. Так как множество коэффициентов STP-фильтра рассчитывается на основе последовательно поступающих фреймов сигнала, то это влечет за собой две потенциальные проблемы, а именно: задержку и неточность (ошибку) их определения. LPC-анализ не может быть завершен до тех пор, пока весь фрейм или обрабатываемый фрагмент не будут доступны для расчета. Следовательно, вносится алгоритмическая задержка как минимум длительностью в один фрейм (обрабатываемый блок). Эта задержка может быть устранена посредством использования LPC-анализа по предыстории, т.е. при использовании только квантованных (или предыдущих) отсчетов для определения LPC-коэффициентов. Однако подобный метод может работать на скоростях значительно больше 10 Кбит/с, так как точность LPC-анализа быстро снижается, что приводит к увеличению шума в реконструированной речи.

LPC-анализ на основе обновления параметров фрейм за фреймом обеспечивает неплохие характеристики в течение длительных фрагментов с медленно изменяющимися спектральными характеристиками. Однако в речевых регионах (фреймах), являющихся наиболее перцептуально важными, обновление параметров LPC от фрейма к фрейму ухудшается, если переходы спадают к концу выделенного для анализа фрейма. В этом случае рассчитанный набор параметров будет представлять среднее значение изменений формы спектральных характеристик в данном сегменте. Для устранения вышеописанных недостатков используется техника блочного LPC-анализа - интерполяции от фрейма к фрейму. Метод интерполяции позволяет достичь улучшения спектрального представления при вычислении промежуточных наборов параметров между фреймами, что обеспечивает более гладкий переход между ними без увеличения сложности, но с внесением малой задержки. Схемы описанных методов интерполяции приведены на рис. 2.2 и 2.3.

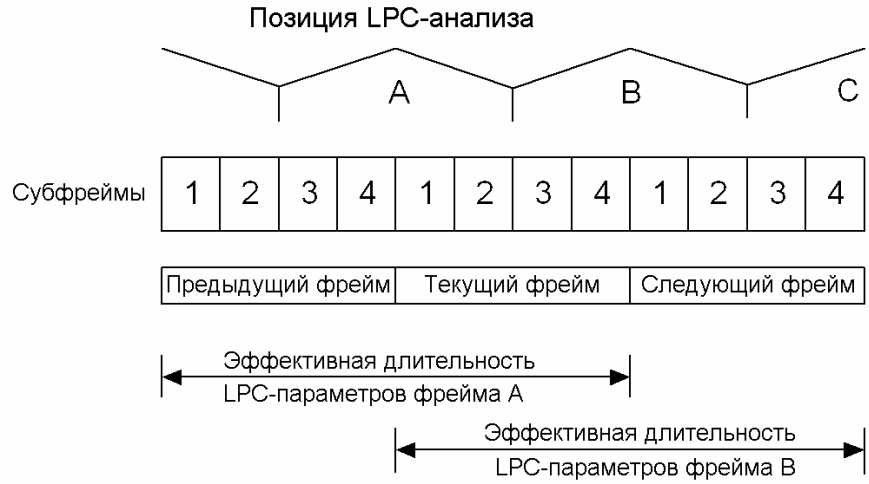


Рис. 2.2. Блочный LPC-анализ с задержкой на полфрейма

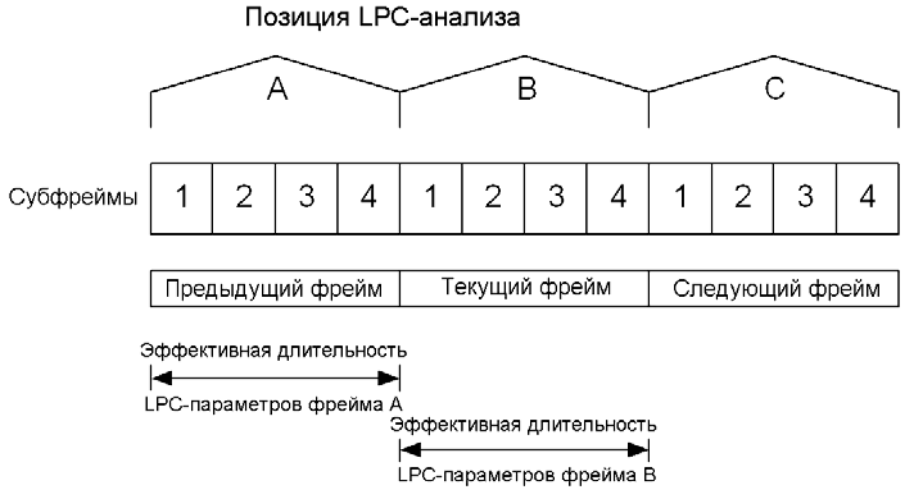


Рис. 2.3. Принцип блочного LPC-анализа без задержки

Спектральная огибающая речевого сегмента длиной L отсчетов может быть аппроксимирована передаточной функцией полюсного фильтра, определяемой по следующему выражению:

$$H(z) = \frac{1}{1 - P_S(z)} = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}}, \qquad (2.1)$$

где $P_s(z)$ – передаточная функция предсказателя. Коэффициенты $\{a_k\}$ определяются на основе метода линейного предсказания (*"Linear Prediction" – LP*). Множество коэффициентов $\{a_k\}$ называется LPC-параметрами или коэффициентами предсказателя. Количество коэффициентов p – порядок предсказателя.

2.1.3. Долговременной фильтр-предсказатель LTP

В то время как STP-фильтр моделирует спектральную огибающую анализируемого речевого сегмента, LTP-фильтр–предсказатель или детектор основного тона – используется для моделирования уточненной структуры этой огибающей. LTP-фильтр очень важен в низкоскоростных кодерах речи, основанных на CELP-алгоритме, где сигнал возбуждения моделируется гауссовым случайным процессом. Поэтому LTP-фильтр используется для обеспечения на выходе фильтра предсказанного сигнала, очень похожего на случайный гауссовый шумовой процесс.

LTP-фильтр имеет малое количество коэффициентов по сравнению с STP-фильтром и описывается следующим выражением:

$$P(z) = 1 - \sum_{i=-I}^{I} \beta_i z^{(-D-i)}, \qquad (2.2)$$

где β_i – коэффициенты усиления фильтра; D – величина задержки в отсчетах.

В CELP-алгоритмах и других схемах «анализа через синтез» LTP-анализ может осуществляться в открытом и замкнутом цикле оптимизации. Описание этих двух методов анализа приводится ниже.

2.2. Линейное предсказание по схеме "анализ через синтез" с перцептуальным взвешивающим фильтром в кодировании речи

Выбор подходящего критерия ошибки в общей модели кодирования речи, представленной на рис. 2.1, является важной составляющей частью проектирования вокодера. Традиционно алгоритмы кодирования речи строятся на основе минимизации среднеквадратического отклонения (СКО) ошибки между оригинальной и синтезированной огибающими речевого сигнала. Однако субъективное восприятие искажений, внесенных в реконструированный сигнал, основывается не только на критерии СКО. Слуховое маскирование показывает, что шум кодирования в формантных областях речи может частично или полностью маскироваться речевым сигналом. Следовательно, большая часть воспринимаемого шума относится к частотным областям с низким (малым) уровнем сигнала. Поэтому для уменьшения степени восприятия внесенного шума в синтезированный речевой сигнал его плоский спектр окрашивается таким образом, чтобы частотные компоненты шумовой составляющей в окрестностях формантных областей имели

более высокую энергию относительно межформантных интервалов. Это можно осуществить с помощью перцептуального взвешивающего фильтра.

Атал и Шрёдер предложили эффективный метод для определения параметров взвешивающего фильтра путем минимизации субъективной громкости шума кодирования, появляющегося в реконструированном речевом сигнале, энергия которого может быть определена согласно следующему выражению:

$$\left|N(f)\right|^2 = \left|\hat{S}(f) - S(f)\right|^2 = \left|\Delta(f)\right|^2 \left|\frac{1-F(f)}{1-P_S(f)}\right|, \tag{2.3}$$

где $|\Delta(f)|^2$ – спектр мощности шума на выходе квантователя; $F(f)$ – фильтр обратной связи; $P_s(f)$ – STP-фильтр.

В модели, представленной на рис. 2.1, взвешивающий фильтр $W'(z)$ может быть выражен следующей передаточной функцией:

$$W'(z) = \frac{1-P_S(z)}{1-F(z)} = \frac{A(z)}{B(z)}, \tag{2.4}$$

которая получена из выражения (2.3) при выполнении замены:

$$\Delta(f) = \left|\hat{S}(f) - S(f)\right| \cdot \frac{1-P_S(f)}{1-F(f)} = N(f)W'(f). \tag{2.5}$$

Наиболее подходящим выбором $B(z)$ является $B(z) = A(z / \gamma)$, который дает следующее выражение передаточной функции перцептуального взвешивающего фильтра:

$$W'(z) = \frac{A(z)}{A\left(z/\gamma\right)} = \frac{1-\sum_{k=1}^{p} a_k z^{-k}}{1-\sum_{k=1}^{p} a_k \gamma^k z^{-k}}, \tag{2.6}$$

где γ – вещественное число от 0 до 1.

Величина коэффициента γ определяется степенью значения формантных областей в спектре ошибки. При этом уменьшение γ увеличивает полосу ω полюсов фильтра $W'(z)$ согласно следующему выражению:

$$\omega = -\frac{f_S}{\pi}\ln(\gamma), \tag{2.7}$$

где f_s – частота дискретизации.

Выбор $\gamma = 0$ дает фильтр с передаточной функцией $W'(z) = A(z)$. В этом случае шум на выходе кодера будет иметь ту же огибающую, что и оригинальный сигнал. С другой стороны, выбор $\gamma = 1$ дает фильтр $W'(z) = 1$, что эквивалентно отсутствию взвешивания. Как показали экспериментальные исследования, для узкополосной речи (50…4000 Гц), оптимальное значение коэффициента γ находится в пределах 0,8…0,9.

Базовая структура кодирования на основе линейного предсказания по схеме "анализ через синтез", содержащая перцептуальные взвешивающие фильтры (2.6),

представлена на рис. 2.4. Здесь осуществляется взвешивание оригинальной речи и синтезированного речевого сигнала отдельно друг от друга перед их вычитанием.

В данной конфигурации кодера синтезирующий STP-фильтр объединен с взвешивающим фильтром. Результирующая передаточная функция синтезирующего перцептуально взвешивающего фильтра $W(z)$ описывается следующим выражением:

$$W(z) = \frac{1}{1 - \sum_{k=1}^{p} a_k \gamma^k z^{-k}}. \qquad (2.8)$$

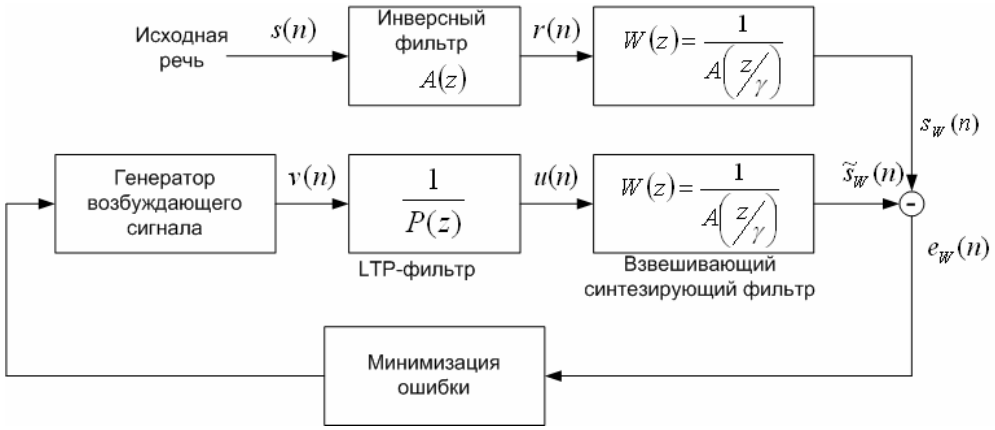


Рис. 2.4. Базовая структура кодирования на основе линейного предсказания по схеме "анализ через синтез"

Для пояснения принципа работы синтезирующего перцептуально взвешивающего фильтра на рис. 2.5 приведен пример АЧХ перцептуально взвешивающего фильтра для различных значений коэффициента γ и огибающая спектра некоторого произвольного фрагмента речевого сигнала (АЧХ STP-фильтра).

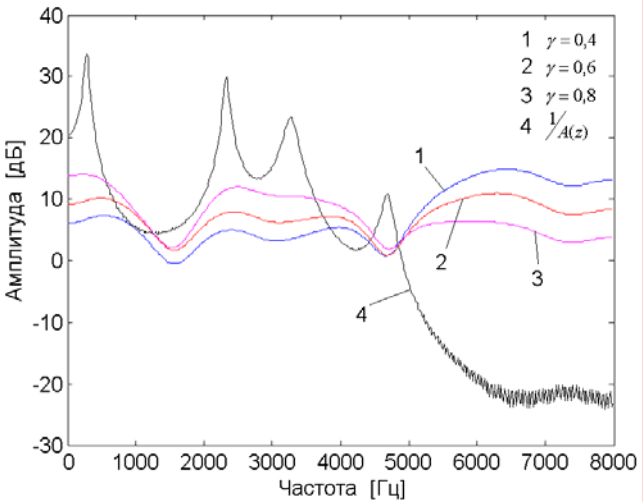


Рис. 2.5. АЧХ синтезирующего и перцептуально взвешивающего фильтров для различных значений коэффициента γ

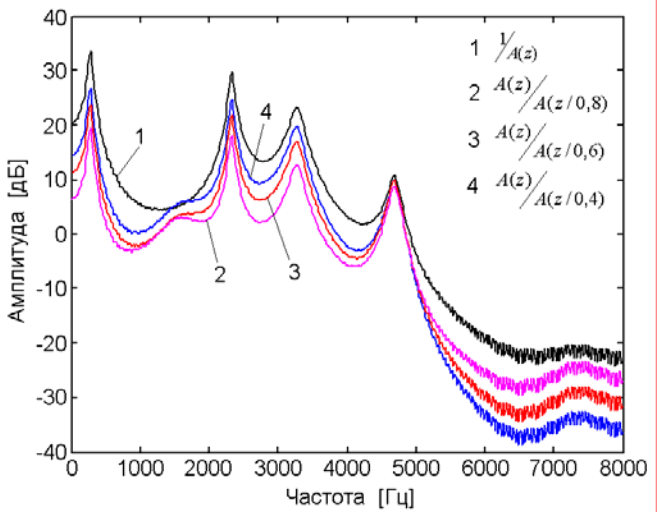


Рис. 2.6. Результирующая АЧХ синтезирующего перцептульно взвешивающего фильтра

Сравнение огибающей спектра некоторого произвольного фрагмента речевого сигнала (АЧХ STP-фильтра, т.е. АЧХ синтезирующего фильтра) с АЧХ

синтезирующего фильтра после перцептуального взвешивания (АЧХ синтезирующего перцептуально взвешивающего фильтра) (рис. 2.6) показывает, что результирующая АЧХ синтезирующего перцептуально взвешивающего фильтра выделяет только полезную составляющую в речевом сигнале, маскирующую шум кодирования.

Основная идея, стоящая за анализом на основе LP, заключается в том, что очередной отсчет речевого сигнала может быть аппроксимирован линейной комбинацией предыдущих отсчетов, т.е.

$$\tilde{s}(n) = \sum_{k=1}^{p} a_k s(n-k), \tag{2.9}$$

где $\tilde{s}(n)$ – предсказанное значение речевого отсчета; $s(n)$ – оригинальный отсчет речевого сигнала.

Таким образом, ошибку предсказания можно представить как разность между оригинальным значением и предсказанным:

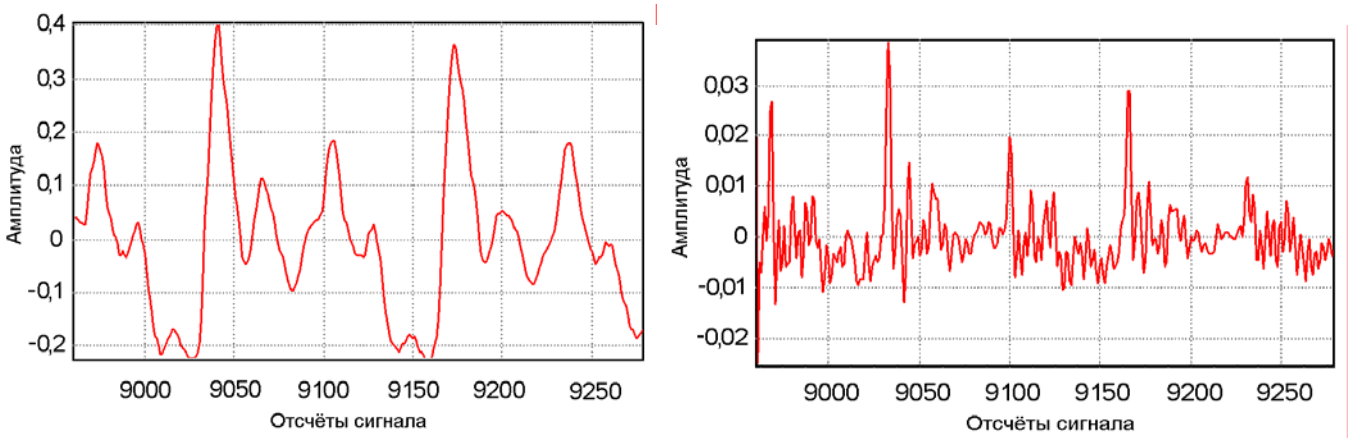
$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^{p} a_k s(n-k). \tag{2.10}$$

Взяв z-преобразование от выражения (2.10), мы получим

$$E(z) = S(z)A(z), \tag{2.11}$$

где $A(z)$ – это инверсия от $H(z)$ в выражении (2.1), т.е. $A(z)$ – инверсный фильтр, передаточная функция которого может быть записана в следующем виде:

$$A(z) = 1 - \sum_{k=1}^{p} a_k z^{-k}. \tag{2.12}$$

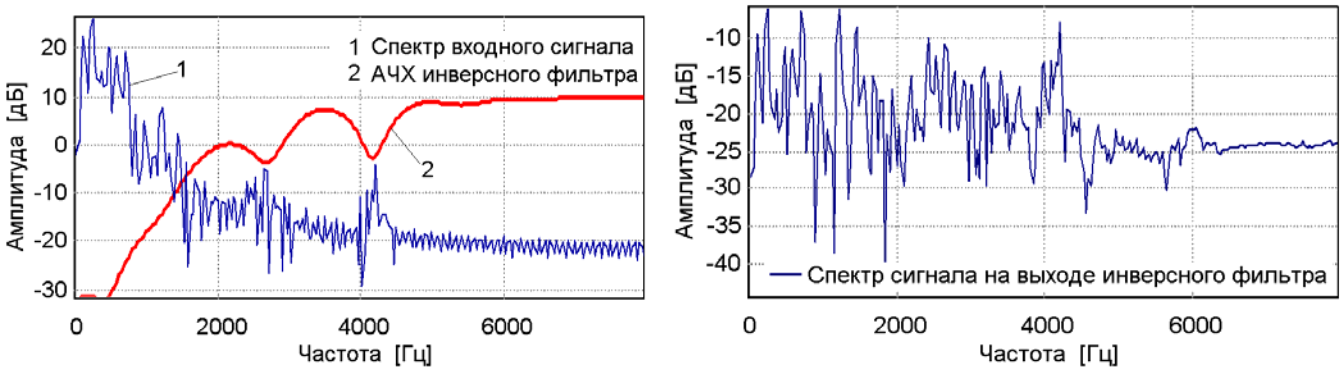


а) б)

Рис. 2.7. К оценке основного тона:
а – входной сигнал; б – остаточный сигнал

Инверсная фильтрация $A(z)$ (см. рис. 2.4) входного речевого сигнала удаляет некоторую избыточность в речи, вычитая из отсчетов оригинального сигнала их предсказанные значения. Остаточный сигнал – сигнал ошибки или сигнал возбуждения очень полезен при оценке основного тона речевого сигнала, так как он

все еще содержит некоторую периодичность (или избыточность), связанную с периодом основного тона оригинальной речи, когда она вокализована. Эта периодичность может иметь частоту повторения 50...500 Гц. Определение частоты основного тона по остаточному сигналу обусловлено тем, что после инверсной фильтрации оригинального речевого сигнала спектральные характеристики сигнала ошибки характеризуются меньшей дисперсией, а спектр остаточного сигнала более плоский. Это иллюстрируется на рис. 2.7 и 2.8 во временной и частотной областях соответственно.



а) б)

Рис. 2.8. К оценке основного тона:
а – спектр входного сигнала; б – спектр остаточного сигнала

2.3. Расчет коэффициентов STP-фильтра

2.3.1. Постановка задачи

Из-за времязависимой природы речевого сигнала коэффициенты предсказания должны определяться на коротких сегментах речи (10...30 мс). При этом необходимо определить множество коэффициентов предсказателя, которые бы минимизировали ошибку на всем сегменте сигнала. Полученные параметры будут являться параметрами системной функции $H(z)$ модели речеобразования в выражении (2.1). Средняя ошибка кратковременного предсказания определяется по следующему выражению:

$$E = \sum_n e^2(n) = \sum_n \left[s(n) - \sum_{k=1}^{p} a_k s(n-k) \right]^2 . \qquad (2.13)$$

Для определения значений $\{a_k\}$, минимизирующих ошибку E, необходимо взять частную производную по всем коэффициентам и приравнять ее к нулю:

$$\frac{\partial E}{\partial a_i} = -2 \sum_n \left\{ \left[s(n) - \sum_{k=1}^{p} a_k s(n-k) \right] s(n-i) \right\} = 0 , \qquad (2.14)$$

что дает следующее уравнение:

$$\sum_n s(n)s(n-i) = \sum_n \sum_{k=1}^{p} a_k s(n-k)s(n-i) . \qquad (2.15)$$

Если изменить порядок суммирования в правой части уравнения (2.15), то получим следующее выражение:

$$\sum_n s(n)s(n-i) = \sum_{k=1}^{p} a_k \sum_n s(n-k)s(n-i)\,,\ i = 1,..., p\,. \tag{2.16}$$

Сделав замену

$$\phi(i,k) = \sum_n s(n-i)s(n-k)\,, \tag{2.17}$$

уравнение (2.16) можем записать в следующем виде:

$$\sum_{k=1}^{p} a_k \phi(i,k) = \phi(i,0)\,,\ i = 1,..., p\,. \tag{2.18}$$

Следовательно, система из p уравнений с p неизвестными может быть эффективно решена относительно неизвестных коэффициентов $\{a_k\}$. В настоящее время применяется множество методов LP-анализа, но наибольшее распространение получили методы на основе автокорреляции и автоковариации, при этом первый имеет меньшую вычислительную сложность и всегда обеспечивает синтез стабильного STP-фильтра.

2.3.2. LP-анализ на основе автокорреляции

В соответствии с формулой (2.13) ошибка определяется на бесконечном интервале $-\infty < n < \infty$. Это не может быть реализовано на практике, поэтому предполагают, что речевой сегмент имеет нулевое значение за пределами интервала $0 \le n \le L-1$ (длины анализируемого фрейма). Это эквивалентно умножению входного сигнала на окно с конечной длиной, отсчеты которого также раны нулю, вне интервала анализа. Учитывая это, выражение (2.17) запишем в следующем виде:

$$\phi(i,k) = \sum_{n=0}^{L+p-1} s(n-i)s(n-k)\,,\ i = 1,..., p\,,\ k = 1,..., p\,. \tag{2.19}$$

Замена $m = n - i$ в формуле (2.19) приводит к следующему выражению:

$$\phi(i,k) = \sum_{n=0}^{L-1-(i-k)} s(m)s(m+i-k)\,. \tag{2.20}$$

Следовательно, $\phi(i,k)$ – это кратковременная автокорреляция $s(m)$, рассчитанная для $(i-k)$.

Пусть $\phi(i,k) = R(i-k)$, где

$$R(j) = \sum_{n=0}^{L-1-j} s(n)s(n+j) = \sum_{n=j}^{L-1} s(n)s(n-j)\,, \tag{2.21}$$

тогда система уравнений (2.18) может быть выражена в следующем виде:

$$\sum_{k=1}^{p} a_k R(|i-k|) = R(i)\,,\ i = 1,..., p\,. \tag{2.22}$$

Или в матричной форме:

$$\begin{pmatrix} R(0) & R(1) & R(2) & \cdots & R(p-1) \\ R(1) & R(0) & R(1) & \cdots & R(p-2) \\ R(2) & R(1) & R(0) & \cdots & R(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R(p-1) & R(p-2) & R(p-3) & \cdots & R(0) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} R(1) \\ R(2) \\ R(3) \\ \vdots \\ R(p) \end{pmatrix}. \quad (2.23)$$

Матрица автокорреляционных значений размером $p \times p$ является симметричной теплицевой матрицей, т.е. все элементы вдоль ее главной диагонали равны. Это свойство может быть использовано для получения эффективного в вычислительном смысле алгоритма для решения данного уравнения известного как алгоритм Дарбина, который формулируется следующим образом:

$E(0) = R(0)$;

for $i = 1$ to p do begin

$$k_i = \left[R(i) - \sum_{j=1}^{i-1} a_j^{(i-1)} R(i-j) \right] / E(i-1); \quad (2.24)$$

$$a_i^{(i)} = k_i;$$

for $j = 1$ to i-1 do begin

$$a_j^{(i)} = a_j^{(i-1)} - k_i a_{i-j}^{(i-1)}; \quad (2.25)$$

$$E(i) = (1 - k_i^2) E(i-1); \quad (2.26)$$

end;

end;

$$a_j = a_j^p, \qquad j = 1,..., p. \quad (2.27)$$

Элемент $E(i)$ в уравнении (2.26) – это ошибка предсказания предсказателя i-го порядка. Промежуточные элементы k_i известны как коэффициенты отражения (коэффициенты модели динамической трубы вокального тракта). Значения коэффициентов k_i находятся в диапазоне $-1 \le k_i \le 1$. Это условие, налагаемое на параметры k_i, является необходимым и достаточным для того, чтобы все корни полинома $A(z)$ находились внутри единичной окружности, что гарантирует стабильность системы $H(z)$.

2.4. Определение параметров LTP

2.4.1. LTP-анализ по методу открытого цикла (OLM)

В методе OLM остаточный сигнал формируется путем фильтрации оригинальной речи инверсным фильтром с LPC-коэффициентами. Затем на каждом субфрейме остаточного сигнала определяются задержка D и коэффициент усиления β_i. На практике задержка D может быть больше, чем длина обновления L LTP-буфера, т.е. $D > L$. Используя этот факт, рассчитанные значения D и β_i получаем более оптимальными. Недостатком при этом является то, что при большом значении L эффективность LTP-фильтра снижается, так как снижается

скорость адаптации задержки D. На рис. 2.9 представлена структура (алгоритм) модифицированного открытого цикла (MOLM) LTP-анализа.

Отставание (запаздывание), или задержка, определяется по выходу LTP-синтезатора, который наиболее близок к оригинальному остаточному сигналу $r(n)$, т.е. для LTP-фильтра первого порядка:

$$e(n) = r(n) - \hat{r}(n) = r(n) - \beta\hat{r}(n-D). \tag{2.28}$$

Следовательно, СКО составит

$$E = \sum_{n=0}^{L-1} e^2(n) = \sum_{n=0}^{L-1} \left[r(n) - \beta\hat{r}(n-D)\right]^2. \tag{2.29}$$

Для определения минимального значения необходимо от выражения E взять первую производную и приравнять ее к нулю, то есть

$$\frac{\partial E}{\partial \beta} = 2\sum_{n=0}^{L-1}\left[r(n) - \beta\hat{r}(n-D)\right] - \hat{r}(n-D) = 0, \tag{2.30}$$

$$\Rightarrow \beta = \frac{\sum_{n=0}^{L-1} r(n)\hat{r}(n-D)}{\sum_{n=0}^{L-1} \hat{r}^2(n-D)}. \tag{2.31}$$

Подставляя оптимальное значение β в выражение (2.29), получим:

$$E = \sum_{n=0}^{L-1} r^2(n) - \frac{\left[\sum_{n=0}^{L-1} r(n)\hat{r}(n-D)\right]^2}{\sum_{n=0}^{L-1} \hat{r}^2(n-D)}. \tag{2.32}$$

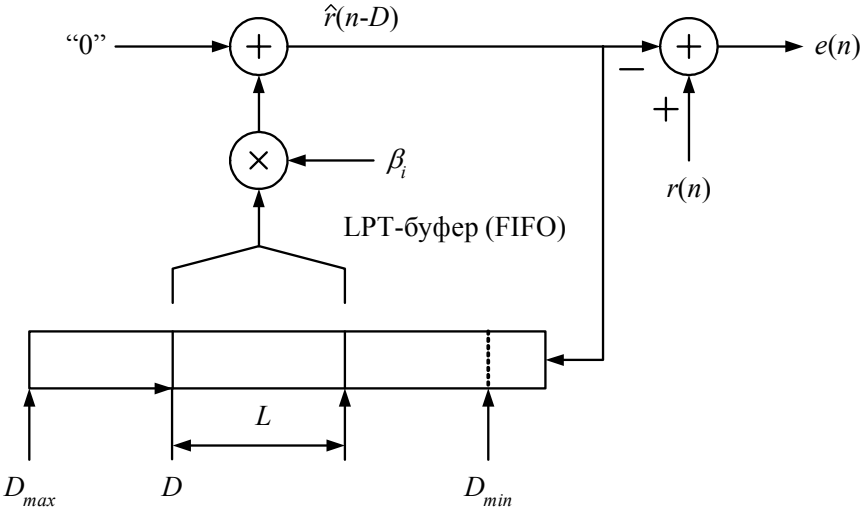


Рис. 2.9. Схема LTP-анализа по методу MOLM

Таким образом, для определения оптимальной задержки D значения запаздывания проверяются для всех значений в диапазоне от $D_{\min}$ до $D_{\max}$. Величина запаздывания, которая минимизирует ошибку E, является оптимальной задержкой

D_{opt}. Подобная формулировка может быть применена и к LTP-фильтру более высокого порядка. Модифицированный метод открытого цикла обеспечивает лучшее качество, чем обычный OLM, так как значения, используемые для определения коэффициентов задержки и усиления, являются квантованными значениями $r(n)$, а не оригинальными. В синтезаторе обеспечивается отсутствие ошибок, связанных с потерей коэффициентов усиления из-за ошибок на предыдущих анализируемых фреймах. MOLM является упрощенной версией схемы определения LTP-параметров по методу замкнутого цикла (CLM) с минимизацией ошибки по выходу синтезирующего STP-фильтра.

2.4.2. Определение параметров LTP по методу замкнутого цикла (CLM)

В системах кодирования по методу "анализ через синтез" наиболее интересным является минимизация ошибки между взвешенным оригиналом и синтезированной выходной речью. Структура процедуры поиска изображена на рис. 2.10. После определения параметров STP остается определить $Gx(n)$, D и β_i. Эти параметры могут быть определены путем полного перебора всех возможных комбинаций; при этом процедура поиска становится очень сложной в вычислительном плане, поэтому на практике применяется субоптимальное решение. Сначала сигнал возбуждения приравнивается к нулю $Gx(n) = 0$ и рассчитываются LTP-параметры, минимизирующие ошибку $e(n)$. Затем LTP-параметры фиксируются и осуществляется поиск сигнала возбуждения и его коэффициента усиления $Gx(n)$.

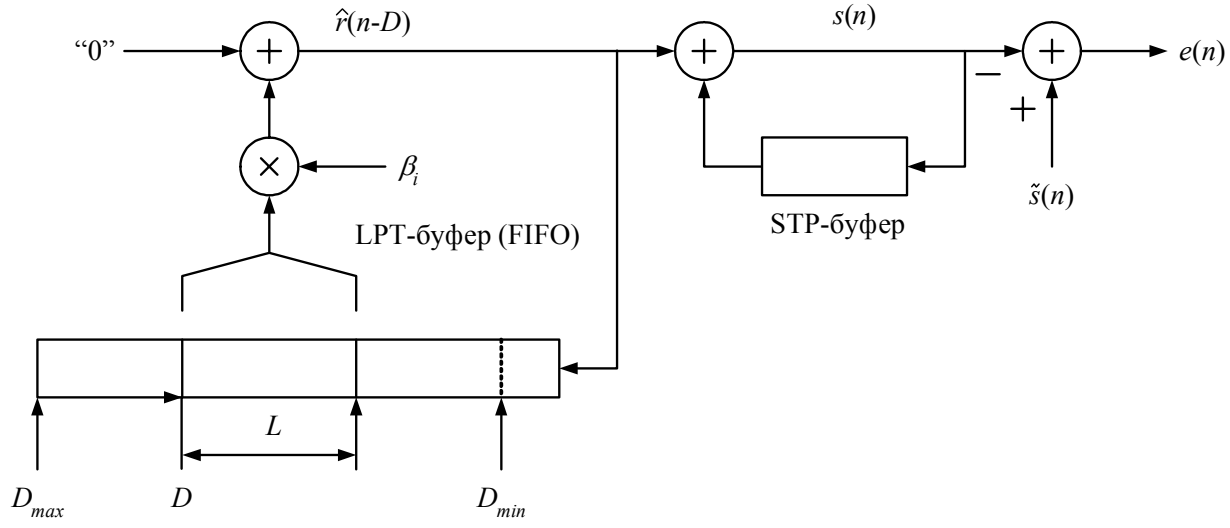


Рис. 2.10. Схема LTP-анализа по методу CLM

Полагая

$$x(n) = 0\,,\ 0 \le n \le L-1\,, \tag{2.33}$$

получим

$$\hat{s}(n) = \sum_{i=-I}^{I} \beta_i \sum_{k=0}^{n} \hat{r}(n-k-D-i) h_w(k)\,, \tag{2.34}$$

где $h_w(k)$ – взвешенная импульсная характеристика STP-фильтра.

Тогда взвешенная СКО E_w будет определяться по следующему выражению:

$$E_w(D) = \sum_{n=0}^{L-1} e^2(n) = \sum_{n=0}^{L-1} \left(\tilde{s}(n) - \hat{s}(n)\right)^2 , \tag{2.35}$$

где

$$\tilde{s}(n) = s_w(n) - s_m(n) . \tag{2.36}$$

Здесь $s_m(n)$ – вклад памяти (или нулевого входного сигнала) STP-фильтра; $s_w(n)$ – взвешенная оригинальная речь.

Следовательно, производную от функции ошибки E_w по коэффициенту усиления β можем записать в виде

$$\frac{\partial E_w}{\partial \beta_i} = 2\left[\sum_{n=0}^{L-1} \tilde{s}(n) - \sum_{i=-I}^{I} \beta_i \sum_{k=0}^{n} \hat{r}(n-k-D-i) h_w(k)\right] \times \left[-\sum_{k=0}^{n} \hat{r}(n-k-D-i) h_w(k)\right] = 0 . \tag{2.37}$$

Обозначим:

$$Z_i(n) = \sum_{k=0}^{n} \hat{r}(n-k-D-i) h_w(k) , \tag{2.38}$$

тогда

$$\sum_{n=0}^{L-1} \tilde{s}(n) Z_j(n) - \left[\sum_{j=-I}^{I} \beta_i \sum_{n=0}^{L-1} Z_i(n) Z_j(n)\right] = 0 , \; -I \le j \le I . \tag{2.39}$$

Следовательно, можем записать уравнения в матричной форме для LTP-фильтра третьего порядка:

$$\begin{bmatrix} \beta_{-1} \\ \beta \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \phi(-1,-1) & \phi(0,-1) & \phi(1,-1) \\ \phi(-1,0) & \phi(0,0) & \phi(1,0) \\ \phi(-1,1) & \phi(0,1) & \phi(1,1) \end{bmatrix}^{-1} \begin{bmatrix} B(-1) \\ B(0) \\ B(1) \end{bmatrix}, \tag{2.40}$$

а для LTP-фильтра первого порядка, где $i = 0$, получим:

$$\beta_0 = \frac{B(0)}{\phi(0,0)} , \tag{2.41}$$

где

$$\phi(i,j) = \sum_{n=0}^{L-1} Z_i(n) Z_j(n) , \tag{2.42}$$

$$B(i) = \sum_{n=0}^{L-1} \tilde{s}(n) Z_i(n) . \tag{2.43}$$

Найденные коэффициенты LTP-фильтра подставляются обратно в выражение (2.35), и задержка D, для которой минимизируется ошибка $E_w(D)$, является оптимальной D_{opt}, соответствующий ей коэффициент усиления является также оптимальным.

Сигнал возбуждения $Gx(n)$ может быть определен при фиксированных D_{opt} и β_{opt}. На практике оптимальная задержка D_{opt} определяется до того, как вычисляется коэффициент усиления LTP-фильтра первого порядка. Уравнения (2.35),(2.41) могут быть записаны следующим образом:

$$E_w(D) = \sum_{n=0}^{L-1}\left[\tilde{s}(n) - \beta_0 \sum_{k=0}^{n} \hat{r}(n-k-D)h_w(k)\right]^2, \tag{2.44}$$

$$\beta_0 = \frac{\sum_{n=0}^{L-1} \tilde{s}(n)Z_D(n)}{\sum_{n=0}^{L-1} Z_D^2(n)}, \tag{2.45}$$

где

$$Z_D(n) = \sum_{k=0}^{n} \hat{r}(n-k-D)h_w(k). \tag{2.46}$$

Подставляя выражения (2.45), (2.46) в (2.44), получим:

$$E_w(D) = \sum_{n=0}^{L-1} \tilde{s}^2(n) - \frac{\left[\sum_{n=0}^{L-1} \tilde{s}(n)Z_D(n)\right]^2}{\sum_{n=0}^{L-1} Z_D^2(n)}. \tag{2.47}$$

На практике второй член выражения (2.47) максимизируется при оптимальном значении D_{opt}, а затем по выражению (2.45) вычисляется коэффициент усиления β_0.

3. ШИРОКОПОЛОСНЫЙ ВОКОДЕР С ПСИХОАКУСТИЧЕСКОЙ МОТИВАЦИЕЙ НА ОСНОВЕ CELP-МОДЕЛИ С МНОГОПОЛОСНЫМ ВОЗБУЖДЕНИЕМ

3.1. Разбиение частотного диапазона на полосы

Человеческое ухо воспринимает звук, границы частот которого задаются в пределах от 20...50 до 22000 Гц. На основании выражения (1.2) можно построить графическую зависимость для широкополосного речевого сигнала 50...8000 Гц (рис. 3.1). Анализ частотной шкалы, представленной на рис. 3.1 и в табл. 1.1, показывает, что весь частотный диапазон 0...8000 Гц может быть аппроксимирован 21 барком и наибольшее количество частотных групп сосредоточено в области низких частот. Данные группы имеют наибольшую перцептуальную важность, так как первые гармоники частоты основного тона в речевых сигналах лежат в пределах 50...500 Гц. На основании графической зависимости (см. рис. 3.1) и данных табл. 1.1 можно сгруппировать критические полосы для получения восьми результирующих полос. Параметры выбранной схемы разбиения частотной шкалы сведены в табл. 3.1.

На основании данных табл. 3.1 может быть спроектирован ДПФ – банк полифазных фильтров (прил. 1). Амплитудно-частотная характеристика такого банка фильтров представлена на рис. 3.2.

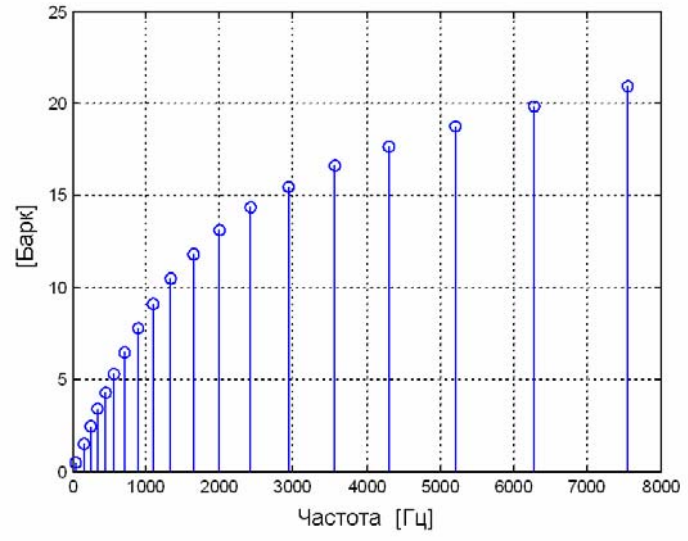


Рис. 3.1. Разбиение частотного диапазона на барки

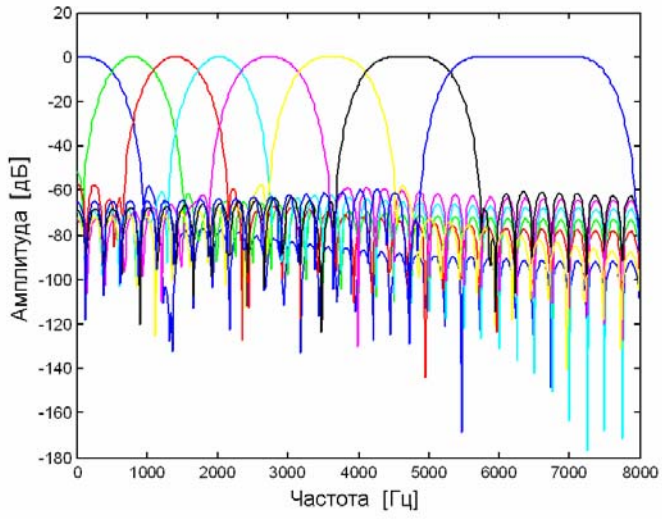


Рис. 3.2. Банк фильтров, осуществляющий разбиение частотной шкалы

Таблица 3.1.

Разбиение частотной шкалы для проектируемого кодера речи

Полоса	Частотный диапазон [Гц]	Ширина полосы [Гц]	Ширина полосы [барк]
1	50…510	460	4
2	510…1080	570	4
3	1080…1720	640	3
4	1720…2320	600	2
5	2320…3150	830	2
6	3150…4100	950	1.5
7	4100…5300	1200	1.5
8	5300…7500	2200	2

3.2. Структура широкополосного CELP-вокодера с многополосным возбуждением

3.2.1. Генерирование кодовых книг

В качестве генератора возбуждающего сигнала в каждой частотной полосе используется кодовая книга: набор полосно-отфильтрованных векторов.

Обучающее множество может состоять из векторов следующих типов:

- гауссовы векторы с единичной дисперсией и нулевым математическим ожиданием, синтезированные генератором случайных чисел;
- векторы, состоящие из отсчетов реальных речевых сигналов нескольких дикторов.

Составление кодовых книг осуществляется для каждой полосы из соответствующего обучающего множества векторов, при этом используется

модифицированный алгоритм *K*-средних. Использование многополосного возбуждения обеспечивает повышение качества речи, обусловленного независимым компонентным анализом, выполняемым в каждой полосе при подборе оптимального сигнала возбуждения. Применение векторного квантования по восьмиполосной кодовой книге входного сигнала проиллюстрировано рис. 3.3.

Рис. 3.3. Векторное квантование входного сигнала

На рис. 3.4 показано квантование вектора входного сигнала в каждой из восьми полос (для большей наглядности представлен неидеальный случай).

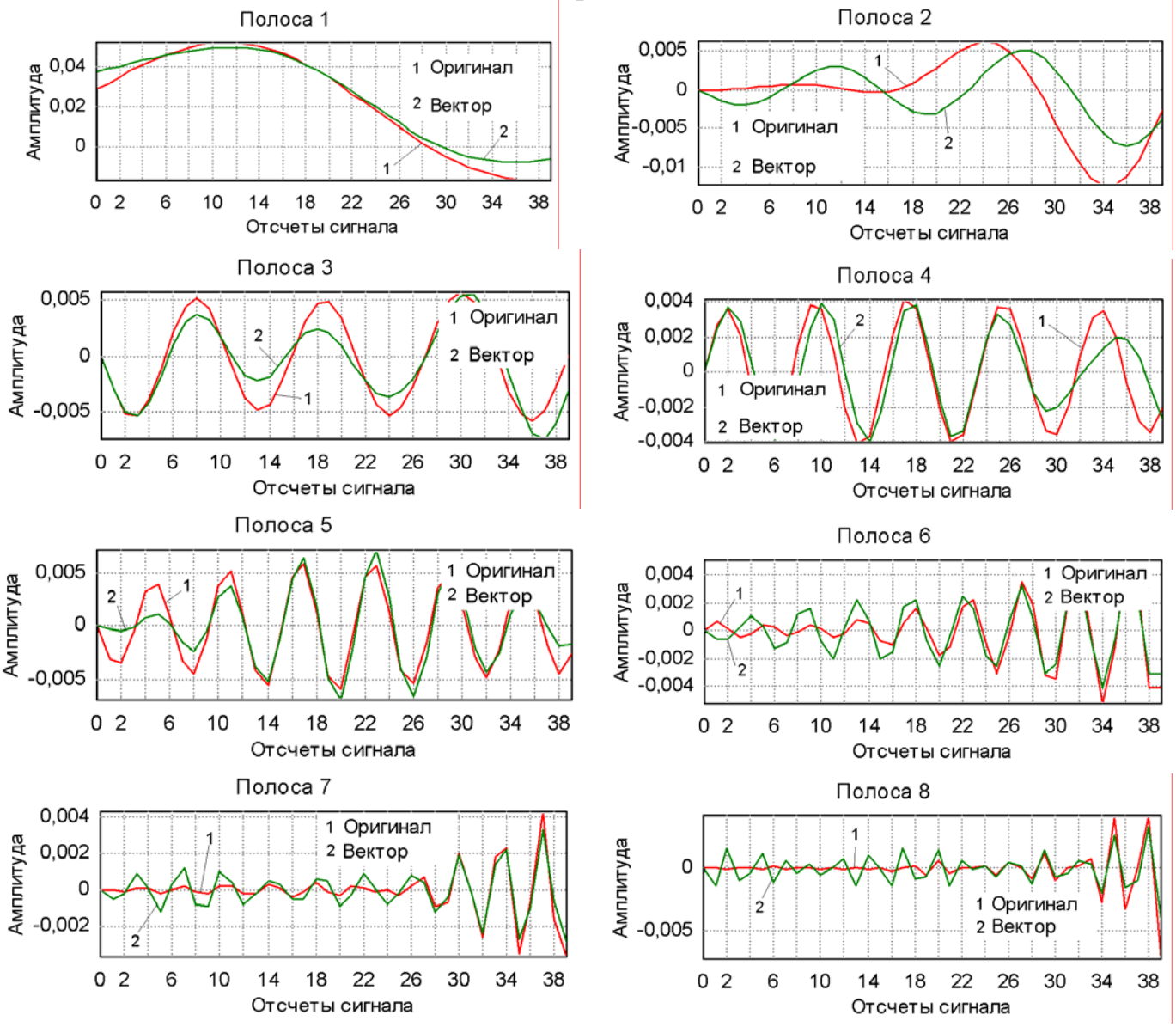


Рис. 3.4. Векторное квантование в критических полосах

Структурная схема широкополосного CELP-вокодера с многополосным возбуждением и психоакустической мотивацией представлена на рис. 3.5. Процесс многоуровневого векторного квантования в структуре данного кодера можно описать в виде следующих выражений:

$$e_{w1}(n) = s_w(n) - G_1\hat{s}_{w1}(n),\ e_{wi}(n) = e_{w(i-1)} - G_i\hat{s}_{wi}(n), 2 \le i \le 8,\ e_w = s(n) - \sum_{i=1}^{8} G_i\hat{s}_{wi}, \qquad (3.1)$$

где $s_w(n)$ – взвешенная оригинальная речь; $\hat{s}_{wi}(n)$ – взвешенная синтезированная речь i-го уровня (субполосы); G_i – коэффициент усиления сигнала i-го уровня (субполосы); $e_{wi}(n)$ - ошибка квантования i-го уровня; $e_{w(i-1)}(n)$ – остаточный сигнал после квантования предыдущего i – 1 уровня (ошибка квантования i – 1 уровня).

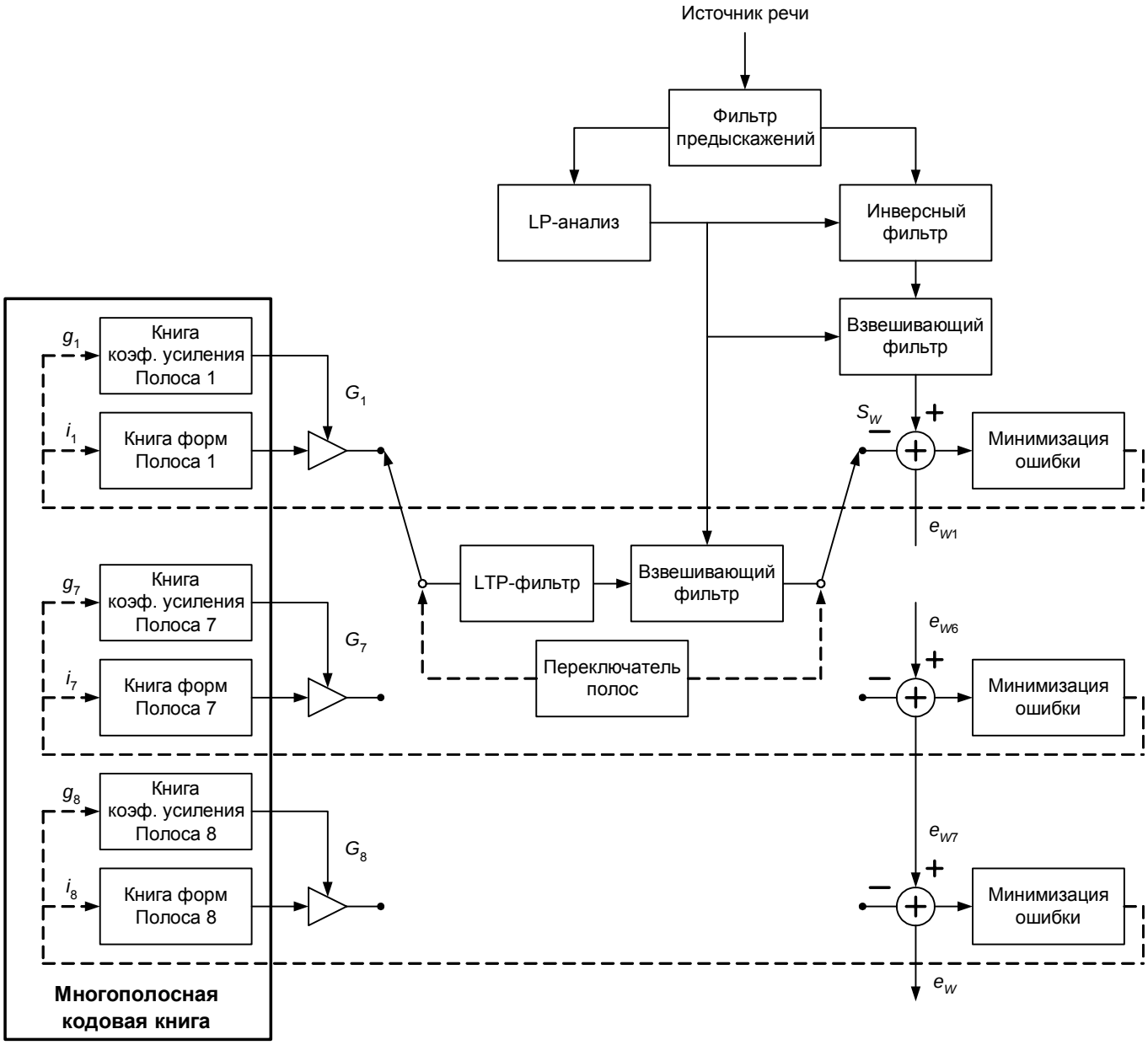


Рис. 3.5.Структура широкополосного CELP-вокодера (кодер)

3.2.2. Поиск параметров модели речеобразования

1. Входной речевой сигнал квантуется при помощи 16 бит на отсчет с частотой дискретизации 16 кГц, после чего кодируется. В широкополосном вокодере используется блочный LP-анализ без интерполяции коэффициентов линейного предсказания.

2. На фрейме входного сигнала длительностью 20 мс (320 отсчетов) осуществляется LP-анализ:

- исходный фрейм взвешивается окном Хэмминга длиной 3·320=960 отсчетов, при этом окно центрируется относительно входного фрейма сигнала;
- осуществляется расчет коэффициентов линейного предсказания и коэффициента усиления модели по описанному выше алгоритму Дарбина. В качестве наиболее оптимального для полосы частот 50…7500 Гц, определен 16-й порядок фильтра-предсказателя (STP).

3. Рассчитанные коэффициенты предсказания используются для получения инверсного фильтра и взвешивающего синтезирующего фильтра.

4. Очередной фрейм входного сигнала фильтруется последовательно соединенными инверсным и взвешивающим синтезирующим фильтрами (коэффициент $\gamma = 0{,}95$), с параметрами, определенными для данного фрейма входного речевого сигнала. Описанные выше фильтры с целью осуществления независимого анализа последовательно поступающих фреймов используются с обнуленной предысторией (без памяти).

5. Взвешенный входной сигнал (фрейм) разбивается на 4 субфрейма длительностью по 5 мс, для каждого из которых осуществляется подбор параметров LTP-фильтра и квантование по восьмиполосной кодовой книге. Параметры LTP-фильтра определяются по взвешенному оригинальному широкополосному сигналу.

6. По методу CLM, описанному выше, осуществляется подбор оптимальных параметров долговременного фильтра предсказателя D_{opt} и β_{opt}, которые фиксируются. При этом оптимальный интервал поиска задержки D составляет 32...150 отсчетов, что соответствует диапазону частоты основного тона 106,(6)...500 Гц. В данном широкополосном кодере используется LTP-фильтр-предсказатель первого порядка. Вокодер имеет восьмиуровневую структуру, т.е. количество уровней соответствует количеству обрабатываемых критических полос.

7. Поиск векторов возбуждения и их оптимальных коэффициентов усиления осуществляется в каждой полосе (на каждом уровне) с последующим вычитанием вклада синтезированного по данному вектору сигнала. Поиск ведется 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 фильтры. После этого из остаточного взвешенного оригинального сигнала вычитается вклад очередной полосы (уровня). Процесс осуществляется последовательно для всех восьми полос.

8. По окончании процедуры поиска выполняется синтез речевого вектора по оптимальному восьмиполосному возбуждающему сигналу, LTP-буфер обновляется синтезированным вектором сигнала, а также осуществляется обновление памяти взвешивающего синтезирующего фильтра.

9. Описанная выше последовательность действий повторяется для всех оставшихся субфреймов. Таким образом, на выходе имеем следующий набор параметров модели речеобразования, которые необходимо передать на сторону декодера: 1) коэффициенты линейного предсказания $\{a_k\}$ – 16 шт.; 2) коэффициент усиления модели G – 1 шт.; 3) задержка LTP-фильтра D_{opt} – 4 шт.; 4) коэффициент усиления LTP-фильтра β_{opt} – 4 шт.; 5) коэффициенты усиления возбуждающих векторов g_i – $8 \cdot 4 = 32$ шт.; 6) индексы векторов возбуждения в каждой из 8 книг i – $8 \cdot 4 = 32$ шт.

3.2.3. Квантование параметров модели речеобразования

Коэффициенты линейного предсказания квантуются на основе метода линейных спектральных пар (частот) (LSF). Схема квантования параметров модели речеобразования представлена в табл. 3.2.

Таблица 3.2.

Параметр	Периодичность обновления на интервале	Кол.	Бит/параметр	Кол-во бит	Интервал, мс	Скорость потока, бит/с
LSF	1	16	3	48	20	2400
Коэфф. усиления модели G	1	1	4	4		200
Коэфф. усиления LTP-фильтра β_{opt}	4	1	4	16		800
Задержка LTP-фильтра D_{opt}	4	1	8	32		1600
Коэфф. усиления сигнала g_i	4	8	4	128		6400
Индекс в книге i	4	8	7	224		11200
Суммарный поток данных, V_{coder}						22600

3.2.4. Структура декодера речевого сигнала

Структура декодера значительно проще по сравнению со структурой кодера представлена на рис. 3.6.

Реконструкция очередного фрейма сигнала представляет собой следующий процесс. Из канала связи в блок декодирования информации поступает пакет данных, который распаковывается. На выходе данного блока декодера получаются следующие параметры:

- коэффициенты линейного предсказания, представленные в виде линейных спектральных пар (LSF), которые преобразуются в LPC-коэффициенты $\{a_k\}$;

- восстанавливается коэффициент усиления модели G;
- задержка LTP-фильтра для каждого субфрейма D_{opt};
- коэффициент усиления LTP-фильтра β_{opt} для каждого субфрейма;
- 8 индексов векторов возбуждения i в соответствующих субполосных кодовых книгах и 8 коэффициентов усиления g_i возбуждающих векторов на каждый субфрейм.

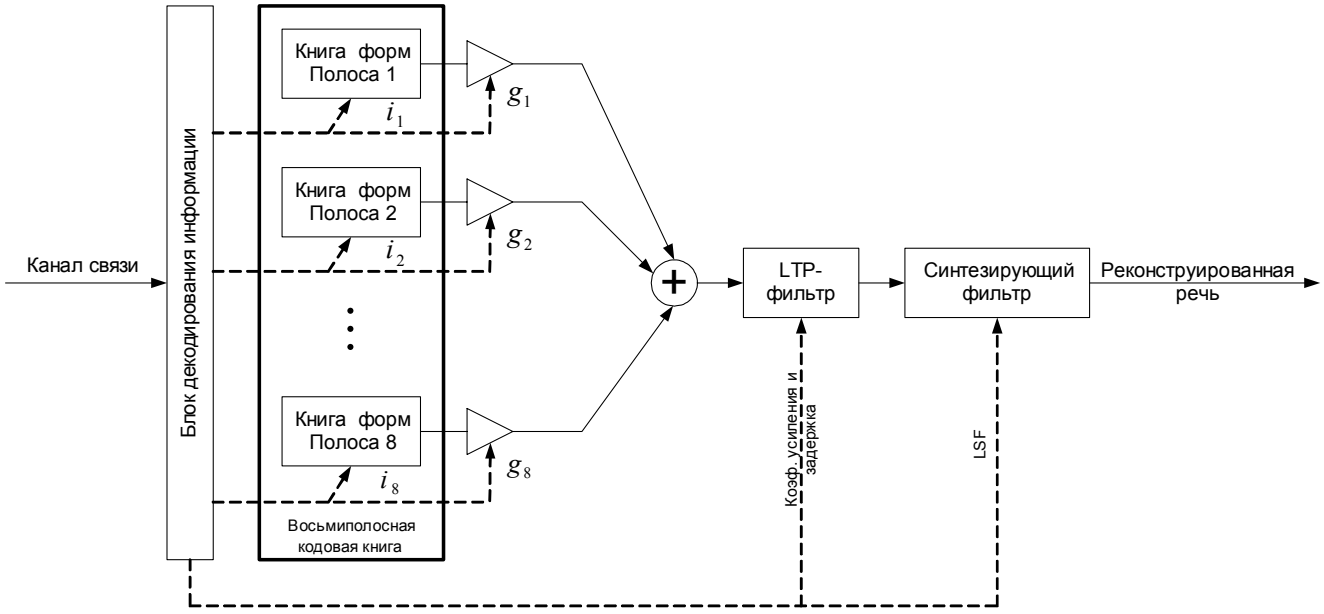


Рис. 3.6. Структура широкополосного CELP-вокодера (декодер)

Для каждого субфрейма по индексам i из кодовых книг восстанавливаются векторы возбуждения, которые усиливаются на соответствующие коэффициенты g_i, после чего суммируются для получения результирующего возбуждающего сигнала, который поступает на LTP-фильтр с восстановленными параметрами D_{opt} и β_{opt}. Выходной сигнал LTP-фильтра синтезируется при помощи STP-фильтра с коэффициентами $\{a_k\}$. В результате на выходе синтезирующего фильтра имеем реконструированную речь. Буфер LTP-фильтра и память синтезирующего фильтра обновляются после синтеза каждого субфрейма.

3.3. Анализ качества широкополосного вокодера

На основании полученных данных о скорости потока можно определить степень компрессии, реализуемую данным кодером. Так как значения отсчетов q входного сигнала квантуются 16 битами, частота дискретизации составляет f_s = 16 кГц, то исходный поток данных определится как:

$$V_{src} = q \cdot f_s = 16 \cdot 16000 = 256 \text{ [Кбит/с]}. \quad (3.1)$$

Тогда степень сжатия составит

$$CMP_{ratio} = \frac{V_{src}}{V_{coder}} = \frac{256}{22.6} = 11.327 \text{ [раз]}. \quad (3.2)$$

Широкополосный компрессор речевого сигнала построен по так называемой «врезной», или встраиваемой, технологии, что выражается в разбиении на восемь

полос и независимом квантовании в каждой их них, т.е. кодер содержит 8 идентичных субкодеров для каждой полосы. Данный принцип позволяет варьировать качество реконструированного речевого сигнала, а следовательно, и скорость потока, в зависимости от требуемой пропускной способности канала связи путем добавления или удаления из рассмотрения высокочастотных полос. Из табл. 3.2 видно, что каждая дополнительная полоса, включенная в рассмотрение (кодирование), вносит вклад в суммарный поток данных, равный 2200 бит/с. Таким образом, обработка только одной самой низкочастотной полосы соответствует минимально возможному для данного кодера потоку данных 7200 бит/с, максимальное качество достигается при обработке всех восьми полос и скорости потока 22600 бит/с.

В табл. 3.3 приведены значения соотношения сигнал/шум на сегменте обрабатываемого сигнала SEGSNR, полученные во время тестирования вокодера. Тестовый материал состоял из четырех фраз (двух женских и двух мужских).

Таблица 3.3

Зависимость соотношения SEGSNR от порядка фильтра-предсказателя и глубины кодовой книги

Порядок STP-фильтра	SEGSNR [дБ]				
	Глубина кодовой книги, уровней				
	64	128	256	512	1024
0	0	0	0	0	0
2	3,3881	4,7633	3,9608	3,7421	5,4044
4	6,0148	7,1172	6,6210	7,4412	8,1268
6	7,2757	7,4497	7,8890	8,6569	8,6699
8	7,9721	8,4349	8,7545	8,7704	9,1823
10	8,1112	8,7190	8,9611	9,2809	10,2173
12	8,6472	8,8144	9,0776	9,4152	10,4642
14	8,8722	9,4011	9,1490	9,8948	10,9460
16	9,0985	9,8644	9,8727	9,9318	11,1049

На последующих рисунках иллюстрируется процесс кодирования речевого сигнала в данном кодере (см. рис. 3.5) на каждом этапе.

Этап 1. Взвешивание исходного речевого фрейма – рис. 3.7.

Рис. 3.7. Фрагменты оригинального и взвешенного оригинального сигнала

Этап 2. Выход LTP-фильтра-предсказателя и траектория частоты основного тона – рис. 3.8 и 3.9 соответственно.

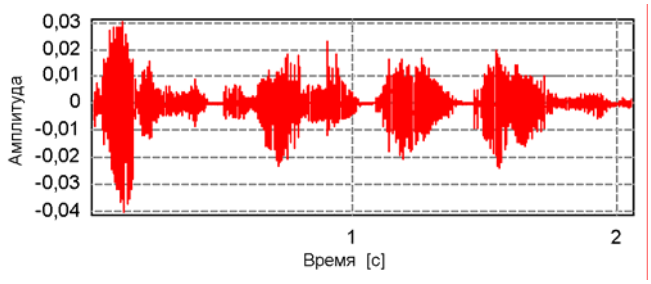


Рис. 3.8. Фрагмент выходного сигнала LTP-фильтра-предсказателя

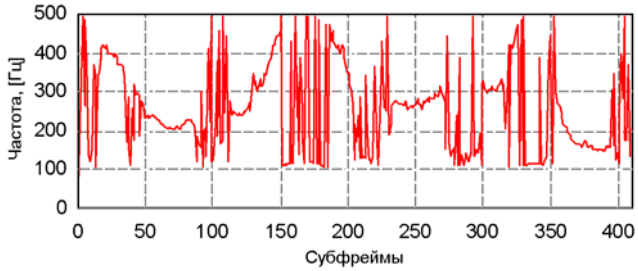


Рис. 3.9. Траектория частоты основного тона

Этап 3. Разностный сигнал (LPC-residual) – рис. 3.10.

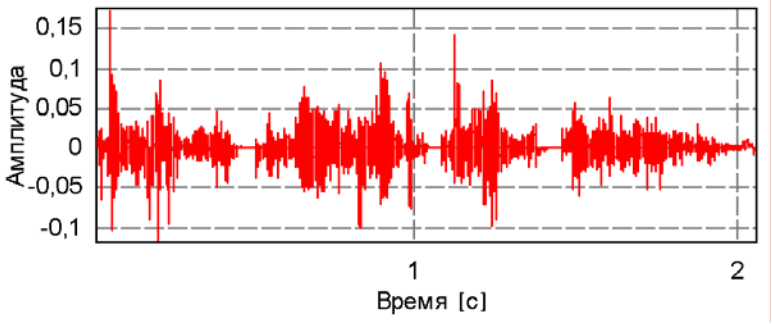
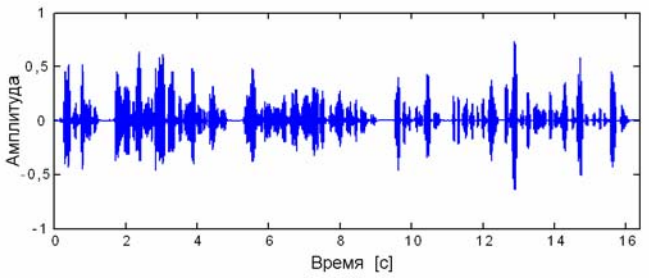
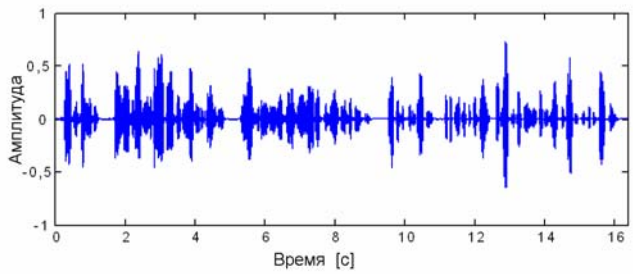


Рис. 3.10. Разностный сигнал между оригинальными и предсказанными значениями речи (фрагмент)

На рис. 3.11 и 3.12 отображены временные фреймы и спектрограммы оригинального и реконструированного речевого сигналов.

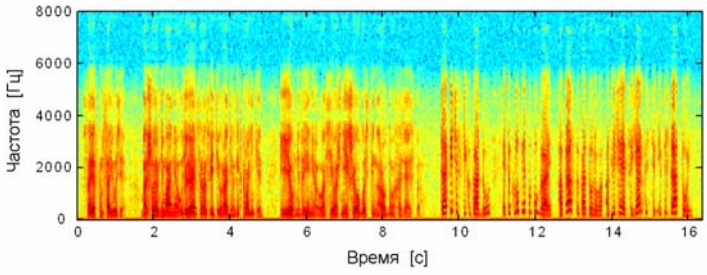


а)

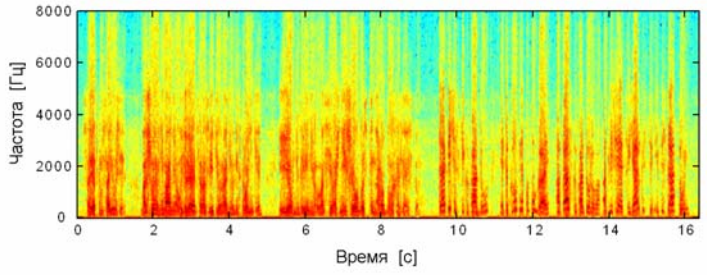


б)

Рис. 3.11. Оригинальный и реконструированный сигнал во временной области:
а – оригинальный сигнал;
б – реконструированный



а)



б)

Рис. 3.12. Спектрограммы оригинального и реконструированного сигналов:
а – оригинальный сигнал;
б – реконструированный

Использование на каждом уровне LTP-фильтра, включенного последовательно с взвешивающим синтезирующим фильтром, позволяет подобрать возбуждающий вектор, наиболее точно повторяющий периодическую структуру (огибающую) оригинального сигнала в соответствующей критической полосе, что вносит существенный вклад в качество реконструированной речи. Единственный недостаток такой модели заключается в увеличении алгоритмической задержки, что связано с поиском параметров модели речеобразования по замкнутому циклу. Однако это наиболее рациональная структура, обеспечивающая высокое качество реконструированной речи. Коэффициенты усиления векторов возбуждения можно также сгруппировать в вектор размерностью 8 и кодировать по книге в 32 уровня (5 бит). Ориентировочная скорость потока составит 15200 бит/с, а степень компрессии увеличится с 11,327 до 16,842. Потенциальные возможности дополнительного применения векторного квантования к параметрам модели речеобразования широкополосного вокодера позволяют снизить скорость потока при незначительном ухудшении качества, но с увеличением алгоритмической задержки.

4. ПЕРЦЕПТУАЛЬНЫЙ ШИРОКОПОЛОСНЫЙ КОДЕР РЕЧИ И АУДИОСИГНАЛОВ НА БАЗЕ ПАКЕТА ДИСКРЕТНОГО ВЭЙВЛЕТ-ПРЕОБРАЗОВАНИЯ (ПДВП)

4.1. Статистическая и перцептуальная избыточность

Основной идеей кодеров является разделение сигнала на частотные компоненты с помощью некоего банка фильтров. Далее компоненты сигнала квантуются в частотной области и общее количество бит динамически распределяется в зависимости от энергии каждого спектрального компонента и его значимости. Пусть в какой-то момент времени спектральные компоненты сигнала обладают одинаковой энергией и занимают весь спектр, а также предполагается отсутствие модуля психоакустического анализа информации. Таким образом, все действия сконцентрированы на устранении статистической избыточности (далее просто избыточности). В данном случае увеличение степени компрессии за счет перераспределения общего количества бит между всеми спектральными компонентами не осуществится в силу того, что для кодирования каждого компонента потребуется одно и то же количество бит. С другой стороны, если допустить, что спектр сигнала «окрашенный», например, основные спектральные компоненты сконцентрированы в области нижних частот, то произойдет перераспределение общего количества бит между всеми спектральными компонентами и значение степени компрессии увеличится. Здесь сигнал содержит избыточность и соответственно в большей или в меньшей степени её можно устранить. Эффективность этой операции зависит от характеристик применяемого банка фильтров.

Пусть x_k – k-я спектральная компонента сигнала, а $Q(x_k)$ – её R_k-битный квантованный аналог, Q – операция квантования, тогда ошибка реконструкции k-й компоненты равна $q_k = x_k - Q^{-1}(Q(x_k))$. Другими словами, q_k – внесенное искажение в

сигнал в результате его кодирования. Среднее число бит на одну спектральную компоненту равно

$$R = \frac{1}{N}\sum_{k=0}^{N-1} R_k \ , \tag{4.1}$$

где N – количество спектральных компонент (каналов в банке фильтров). Принимая во внимание, что шум квантователя является белым, дисперсия внесенных искажений в сигнал в результате кодирования для ИКМ квантователя равна

$$q^2 = \frac{1}{N}\sum_{k=0}^{N-1}\left(\frac{x_k^2}{3\cdot 2^{2R_k}}\right). \tag{4.2}$$

Целью оптимизации является минимизация дисперсии ошибок реконструкции q^2 при ограничении на общее распределение бит. Число уровней реконструкции для квантования компоненты *k*-го канала банка фильтров $L_k = 2^{R_k}$, тогда

$$R = \frac{1}{N}\sum_{k=0}^{N-1}\log_2 L_k = \frac{1}{N}\log_2\prod_{k=0}^{N-1} L_k \ . \tag{4.3}$$

Далее

$$2^R = \prod_{k=0}^{N-1} L_k = L_g^N \text{, где } L_g = \left(\prod_{k=0}^{N-1} L_k\right)^{\frac{1}{N}} \tag{4.4}$$

является средним геометрическим значением уровней реконструкции квантователя. Минимизация дисперсии внесенных искажений при кодировании сигнала основывается на методе множителей Лагранжа λ:

$$\frac{d}{dL_k}\left\{\frac{1}{N}\sum_{k=0}^{N-1}\frac{x_k^2}{3\cdot L_k^2} + \lambda\prod_{k=0}^{N-1} L_k\right\} = 0 \, . \tag{4.5}$$

После дифференцирования и некоторых преобразований формула оптимального распределения бит по каналам банка фильтров примет вид

$$R_k = R + \frac{1}{2}\log_2\left(x_k^2\right) - \frac{1}{2}\log_2\left(\prod_{k=0}^{N-1} x_k^2\right)^{\frac{1}{N}} . \tag{4.6}$$

Из выражения (4.6) следует, что минимальное число бит в каждом *k*-м канале определяется распределением спектральной энергии в сигнале и выигрыш в количестве бит по сравнению с однополосным банком фильтров будет только в том случае, когда среднегеометрическое значение спектральной плотности мощности сигнала будет много меньше её среднеарифметического значения. Отношение среднегеометрического значения спектральной плотности мощности сигнала к её среднеарифметическому значению есть мера пологости спектра сигнала (*Spectral Flatness Measure – SFM*):

$$SFM = \frac{\left(\prod_{k=0}^{N-1} x_k^2\right)^{\frac{1}{N}}}{\frac{1}{N}\sum_{k=0}^{N-1} x_k^2}. \qquad (4.7)$$

Из формулы (4.7) видно, что значения *SFM* варьируются в пределе от 0 до 1. Если $SFM = 1$, то подразумевается, что входной сигнал с пологим спектром и соответственно никакого увеличения компрессии нельзя получить. Пусть $SMF = 1$, тогда согласно (4.6) получается, что $R_k = R$. Следует отметить, что *SFM* зависит не только от распределения спектральной энергии сигнала, но также и от разрешающей способности банка фильтров, т.е. от общего числа *N* каналов в банке фильтров.

Таким образом, мерой избыточности в сигнале является мера пологости спектра *SFM*: чем более пологий спектр сигнала, тем меньше избыточности в сигнале. Малое значение *SFM* подразумевает потенциально высокую степень компрессии сигнала, которую, естественно можно, оценить числом бит, необходимых для кодирования сигнала без артефактов. Из вышеприведенных формул может быть получено выражение, показывающее уменьшение энтропии входного сигнала за счет его разбиения банком фильтров.

В перцептуальном кодере сигналов стоит цель не только устранения информационной избыточности, но и изоляции перцептуальной избыточности акустической информации в сигнале. Это желание расположить вносимые искажения, в результате кодирования, в спектре реконструированного сигнала ниже порога маскирования, т.е. порога восприятия акустической информации слушателем. Соотношение сигнал/шум *SNR* для квантования компонент каналов банка фильтров равно

$$SNR = 10\log\frac{x^2}{q^2}, \qquad (4.8)$$

а соотношение сигнал/порог маскирования *T SMR* определяется следующим образом

$$SMR = 10\log\frac{x^2}{T^2}. \qquad (4.9)$$

Далее для компонент сигнала *k*-го канала, значения которых больше порога маскирования T_k, хотелось бы максимизировать разность *SNR*-*SMR*, или, что эквивалентно, минимизировать разность *SMR*-*SNR*. Для соотношения *SMR/SNR* с учетом дисперсии q^2 (4.2) дисперсия внесенных искажений кодированием, взвешенная маскирующим фактором, равна

$$\frac{q^2}{T^2} = \frac{1}{N}\sum_{k=0}^{N-1}\left(\frac{x_k^2 \big/ T_k^2}{3\cdot 2^{2R_k}}\right), \qquad (4.10)$$

где T_k – уровень порога маскирования в *k*-м канале банка фильтров.

Минимизация данной взвешенной ошибки (4.10), аналогично варианту минимизации дисперсии ошибки реконструкции q^2 (4.2), приводит к следующей формуле оптимального распределения бит по каналам банка фильтров:

$$R_k = R + \frac{1}{2}\log_2\left(x_k^2 / T_k^2\right) - \frac{1}{2}\log_2\left(\prod_{k=0}^{N-1} x_k^2 / T_k^2\right)^{\frac{1}{N}}. \quad (4.11)$$

Из формулы (4.11) следует, что мера перцептуальной избыточности определяется как отношение

$$PSFM = \frac{\left(\prod_{k=0}^{N-1} x_k^2 / T_k^2\right)^{\frac{1}{N}}}{\frac{1}{N}\sum_{k=0}^{N-1} x_k^2 / T_k^2}. \quad (4.12)$$

Как видно, *PSFM* зависит от распределения по частотному диапазону спектральной энергии взвешенной энергией порога маскирования. В данном случае необходимо построить частотно-временное преобразование, характеристики которого зависят от временных изменений сигнала, т.е. обеспечивается требуемое разрешение не только по частоте, но и по времени. Характеристика информационной емкости сигнала в частотной области и область его эффективного кодирования схематически показаны на рис. 4.1.

Рис. 4.1. Характеристика информационной емкости аудиосигнала в частотной области и область его эффективного кодирования (стрелками показаны направления уменьшения избыточности соответственно статистической и перцептуальной)

4.2. Общая структура перцептуального кодера

Ключевая концепция кодирования аудиосигналов на основе восприятия акустической информации человеком (перцептуальное кодирование) базируется на так называемом пороге едва различимых искажений, который является функцией спектра входного сигнала и параметров психоакустической модели, а минимальное

число бит, необходимое для кодирования аудиосигнала, оценивается «перцептуальной энтропией» (*PE*):

$$PE = \frac{1}{N}\sum_{f=f_l}^{f_h} \max\left(0, \log_2 \frac{|signal(f)|}{threshold(f)}\right), \tag{4.13}$$

где N – число частотных компонент в частотном диапазоне f_l и f_h; f_l – нижняя частота (например, $f_l = 0$ Гц) диапазона; f_h – верхняя частота (например, $f_h = 22050$ Гц) диапазона; $|signal(f)|$ – амплитуда частотной компоненты f; *threshold(f)* – оценка порога маскирования на частоте f.

На практике, *PE* часто называют функцией Джонстона (Johnston, 1988) и вычисляют на основе полосового анализа аудиосигнала:

$$PE = \sum_{i=1}^{25}\sum_{\omega=bl_i}^{bh_i} \log_2\left(2\left|n\mathrm{int}\left(\frac{\mathrm{Re}(\omega)}{\sqrt{6T_i/k_i}}\right)\right|+1\right) + \log_2\left(2\left|n\mathrm{int}\left(\frac{\mathrm{Im}(\omega)}{\sqrt{6T_i/k_i}}\right)\right|+1\right), \tag{4.14}$$

где i – индекс критической полосы; bl_i и bh_i – нижнее и верхнее значение частоты i-й критической полосы; k_i – количество компонент преобразования в i-й критической полосе; T_i – значение порога маскирования в критической полосе i; $n$int – операция округления до ближайшего целого значения.

Следовательно, та часть сигнала, которая может быть изменена (в общем случае отброшена) и при этом не вносится дополнительных искажений при его восстановлении, является перцептуально избыточной, а та часть сигнала, отражающая слышимую человеком акустическую информацию, измеряется и кодируется.

Структуры большинства кодеров сигналов на основе психоакустики сходны и могут быть представлены обобщенной схемой (рис. 4.2).

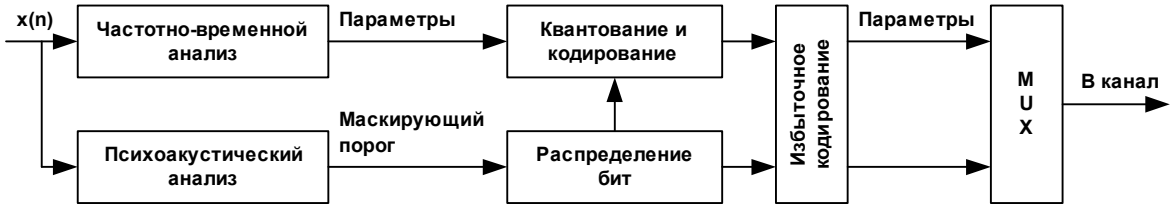


Рис. 4.2. Общая структура перцептуального аудиокодера

Входной аудиосигнал разбивается на квазистационарные фреймы длительностью от 2 до 50 мс в зависимости от алгоритмов обработки и методов кодирования. Блок частотно-временного анализа аппроксимирует временные и спектральные параметры аудиосигнала для каждого анализируемого фрейма с учетом шкалы критических частотных полос. В блоке психоакустического анализа оценивается энергия маскирующего сигнала (уровни маскирующих порогов) на базе психоакустической модели. При этом определяются максимальные искажения, возникающие в каждой точке частотно-временной плоскости в процессе квантования и кодирования частотно-временных оценок без введения искусственного артефакта слышимости при восстановлении сигнала. Следовательно, психоакустический анализатор вычисляет частотно-временной

параметр невосприятия акустической информации слушателем, который затем передается в блок квантования и кодирования. Таким образом, в процессе психоакустического кодирования необходимо: во-первых, установить вид маскирующего сигнала, во-вторых, вычислить соответствующие пороги. Затем полученную информацию следует использовать для того, чтобы расположить спектр шума кодирования ниже так называемого порога едва различимых искажений JND (just noticeable distortion).

4.3. Широкополосный перцептуальный ПДВП-кодер

В настоящем разделе методического пособия исследуется построение перцептуальных кодеров сигналов на фиксированной структуре дерева ПДВП, согласованной со шкалой критических частот восприятия акустической информации человеком (широкополосные кодеры речи). В приложении 2 приводятся краткие сведения о новом теоретическом направлении цифровой обработки сигналов – дискретном вэйвлет-преобразовании.

Структура широкополосного перцептуального ПДВП-кодера речи и аудиосигналов, ядром которой является пакет дискретного вэйвлет-преобразования, представлена на рис. 4.3.

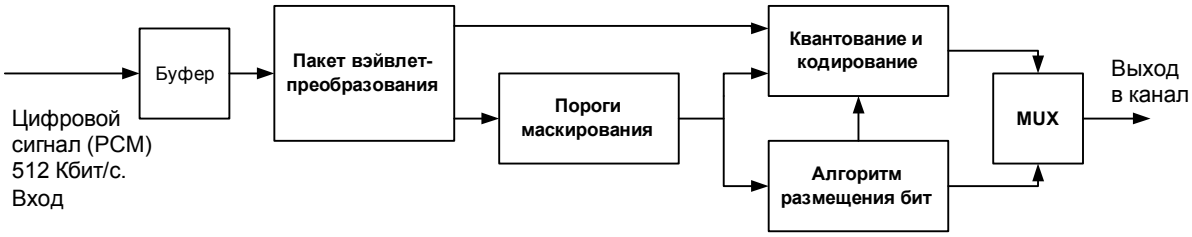


Рис. 4.3. Структура кодера аудиосигналов на базе ПДВП

Обработка входного сигнала выполняется согласно следующей последовательности шагов:

1. Входной сигнал $x(n)$ разбивается на фреймы длиной по 256 отсчётов с перекрытием в 1/16 длины фрейма. Для устранения фазовых разрывов между последовательными фреймами сигнала выполняется взвешивание отсчетов фреймов временным окном Хеннинга.

2. Сегментированный входной сигнал обрабатывается анализирующим банком цифровых фильтров ПДВП, представляющих собой древовидную структуру, состоящую из рекурсивно повторяющихся базовых декомпозиций для анализа входного сигнала (рис. 4.4).

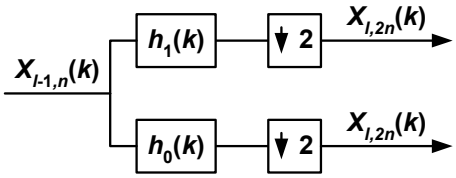


Рис. 4.4. Базовая декомпозиция анализа ПДВП

Выражения, описывающие анализирующую базовую декомпозицию, представлены уравнениями:

$$X_{l,2n}(k)=\sum_{m\in Z}h_0(m-2k)X_{l-1,n}(k) \text{ и } X_{l,2n+1}(k)=\sum_{m\in Z}h_1(m-2k)X_{l-1,n}(k). \quad (4.15)$$

где $h_1(k)$ и $h_0(k)$ – коэффициенты соответственно анализирующего высокочастотного и низкочастотного вэйвлет-фильтров.

ПДВП аппроксимирует критическую шкалу частот таким образом, чтобы расстояние между центральными частотами $z(f)$ полос пропускания было размером в один барк. На рис. 4.5 показана структура дерева ПДВП *CB-WPD*, согласованного с критической шкалой частот восприятия акустической информации человеком (1.2),(1.3):

$$CB\text{-}WPD\text{: } (l,n) \in E_{CB},\ l=\overline{0,6}, \quad (4.16)$$

где E_{CB} – обозначает множество узлов дерева ПДВП, соответствующего *CB-WPD*.

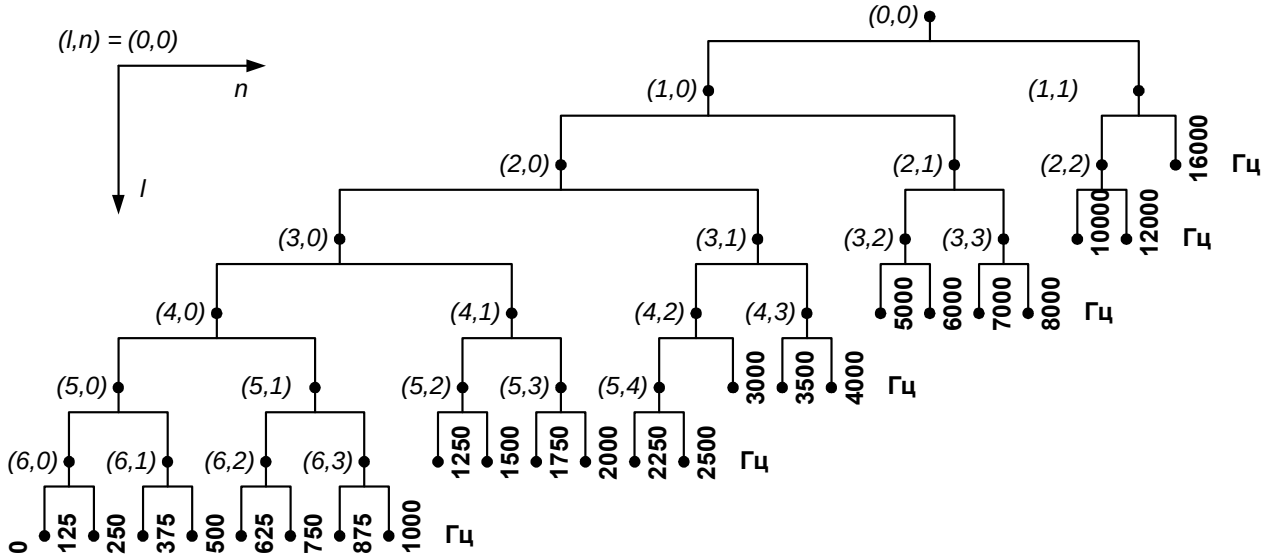
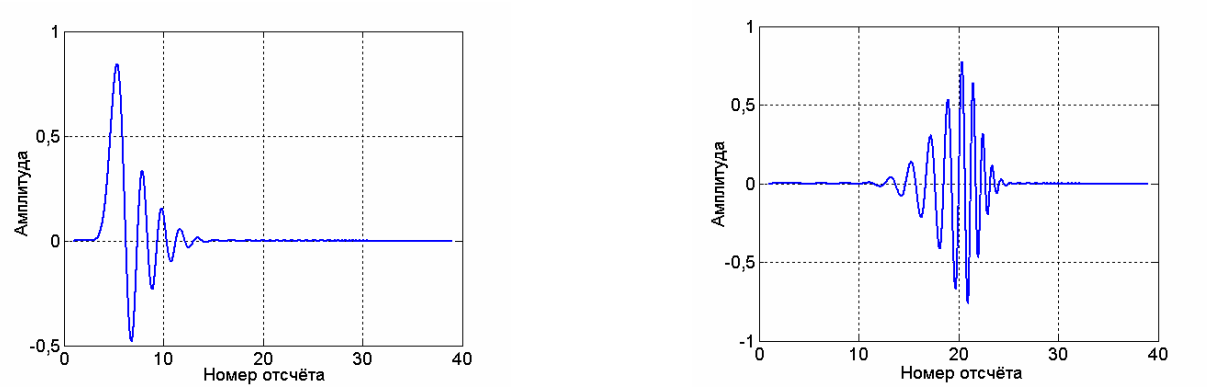


Рис. 4.5. ПДВП $(l,n) \in E_{CB}$

Дерево *CB-WPD* делит частотный диапазон [0...16 кГц] на 24 неравномерных полосы *CBW*(*f*), т.е. на 24 барка. Корневой узел $(l,n) = (0,0)$ данного дерева соответствует всему частотному диапазону сигнала. Каждый внутренний узел дерева $(l,n) \in E$, названный узлом предка, делится на два потомка: 1-й потомок и 2-й потомок, ассоциируемые соответственно с высокочастотной и низкочастотной фильтрацией, выходные сигналы (вэйвлет-коэффициенты) которых децимируются в соотношении 2:1. Вэйвлет-функция семейства Добеши 20-го порядка, обладающая хорошей частотной избирательностью, применяется для анализа входных фреймов (рис. 4.6).



а) б)

Рис. 4.6. Функция Добеши 20-го порядка: а – масштабная; б – вэйвлет-функция

Амплитудно-частотная характеристика (АЧХ) анализирующего банка фильтров структуры ПДВП с вэйвлет-функцией 20-го порядка показана на рис. 4.7.

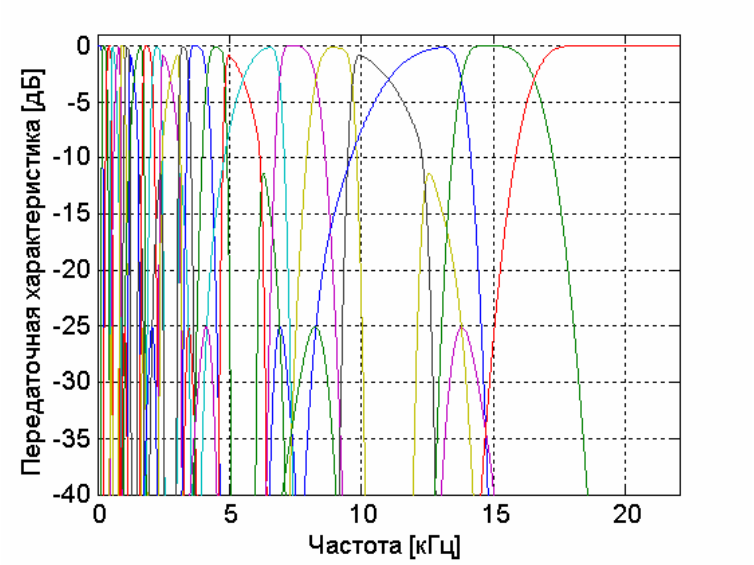


Рис. 4.7. АЧХ банка вэйвлет-фильтров для дерева *CB-WPD*
(вэйвлет-фильтр – Добеши, 20-го порядка)

Частотно-временной план дерева (см. рис.4.5) для ПДВП, согласованного с критической шкалой частот, представлен на рис. 4.8. Ширина каждой клеточки есть длина фрейма и определяется как $F_l = 2^l$ ($F_{min} = 4$ отсчетам и $F_{max} = 128$ отсчетам). Длина анализируемого окна равна $W = (P\text{-}1)(F_l\text{-}1) + 1$ отсчет. Для первого уровня $l = 2$ преобразования определяющей является область верхних частот и длина окна $W = 118$ отсчетам при длине фильтра прототипа $P = 40$. Для уровня $l = 7$ преобразования наибольшая частотная разрешающая способность в области нижних частот, а окно $W = 4952$ отсчетам. Места расположения каждого коэффициента ПДВП $X_{l,n,k}$, где l - номер масштабного уровня преобразования ($0 \le l \le 7$), n - номер узла масштабного уровня преобразования, K – количество вэйвлет-коэффициентов в полосе, показаны в табл. 4.1 и на рис. 4.8.

3. Вычисление порогов маскирования для каждого анализируемого фрейма входного сигнала выполняется в вэйвлет-области согласно процедуре (см. подразд. 4.5).

4. В блоке квантования и кодирования вычисляется шаг квантования для каждой частотной полосы на основе результирующего порога маскирования согласно выражению

$$\delta_{CB}(z) = \sqrt{12 \cdot \frac{T_{CB}(z)}{K(z)}}, \tag{4.17}$$

где $T_{CB}(z)$ – результирующий порог маскирования в полосе z; $K(z)$ – число вэйвлет-коэффициентов в полосе z.

Вэйвлет-коэффициенты квантуются линейным квантователем в каждой полосе с шагом квантования $\delta_{CB}(z)$:

$$qL_{z,k} = \left| n \operatorname{int}\left(\frac{X_{z,k}}{\delta_{CB}(z)} \right) \right| + 0{,}5 . \tag{4.18}$$

Блок «Алгоритм размещения бит» предназначен для определения среднего числа бит на отсчёт входной последовательности, в которой учитывается неперфективность частотно-временного преобразования ПДВП.

Таблица 4.1

№ полосы	Узел		Количество вэйвлет-коэффициентов	Параметры полосы [Гц]		
z	l	n	K	нижняя	центр	верхняя
1	7	0	2	0	62,5	125
2	7	1	2	125	187,5	250
3	7	2	2	250	312,5	375
4	7	3	2	375	437,5	500
5	7	4	2	500	562,5	625
6	7	5	2	625	687,5	750
7	7	6	2	750	812,5	875
8	7	7	2	875	937,5	1000
9	6	4	4	1000	1125	1250
10	6	5	4	1250	1375	1500
11	6	6	4	1500	1625	1750
12	6	7	4	1750	1875	2000
13	6	8	4	2000	2125	2250
14	6	9	4	2250	2375	2500
15	5	5	8	2500	2750	3000
16	5	6	8	3000	3250	3500
17	5	7	8	3500	3750	4000
18	4	4	16	4000	4500	5000
19	4	5	16	5000	5500	6000
20	4	6	16	6000	6500	7000
21	4	7	16	7000	7500	8000
22	3	4	32	8000	9000	10000
23	3	5	32	10000	11000	12000
24	2	3	64	12000	140000	16000

5. Кодирование без потерь заквантованных коэффициентов выполняется на основе кодовых книг Хаффмана: каждому уровню квантования ставится в соответствие из кодовой книги бинарный вектор $B_{z,k}$ длиной w_k.

$$B_{z,k} = \left(b_{k,1}, b_{k,2}, b_{k,3}, \ldots, b_{k,w_k}\right),\ b_{k,j} \in \{0,1\},\ j = \overline{1, w_k}\ . \tag{4.19}$$

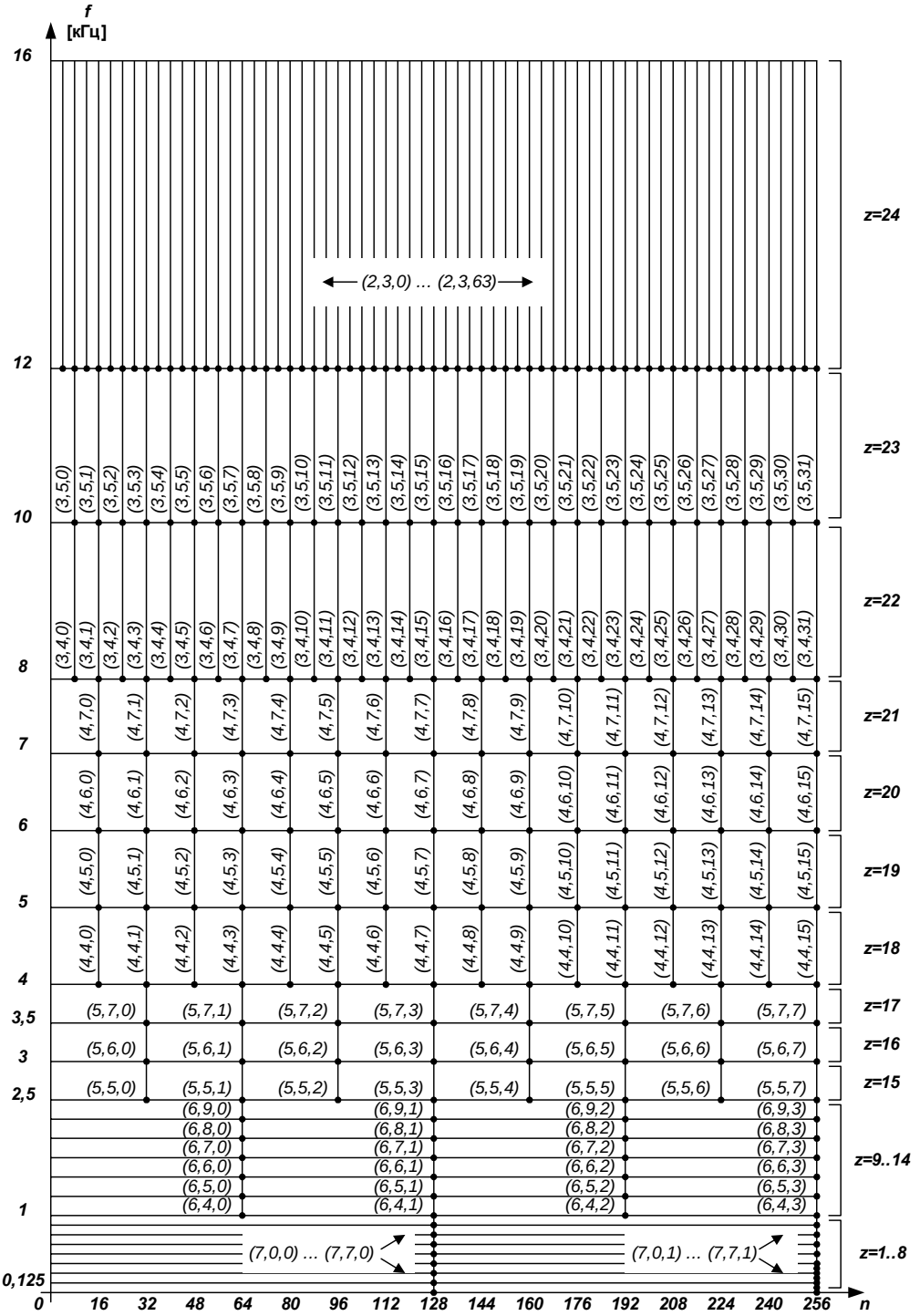


Рис. 4.8. Карта частотно-временного разрешения дерева ПДВП

Причём, очевидно, вероятность данного вектора равна вероятности уровня квантования $p(B_{z,k}) = p(qL_{z,k})$. При этом среднее значение длины кодового слова должно быть не меньше значения энтропии входной последовательности

$$R_z \geq H\left(p\left(qL_{z,k}\right)\right) = \sum_{\substack{k \\ \text{для} \\ \forall(z)\in E_i}} p\left(qL_{zk}\right)\cdot \log_2\left(\frac{1}{p\left(qL_{z,k}\right)}\right). \qquad (4.20)$$

Близость среднего значения длины кодового слова R_z к значению энтропии $H(p(qL_z))$ в (4.20) говорит о приближении к оптимальным кодам представления входных данных. Кодирование будет оптимальным только тогда, когда значение средней длины кодового слова будет равно значению энтропии ($R_z = H(p(qL_z))$, $(z) \in E_i$). Избыточность кодового представления вэйвлет-коэффициентов по Хаффману оценивается согласно формуле

$$cr_z = \frac{R_z - H\left(p\left(qL_z\right)\right)}{H\left(p\left(qL_z\right)\right)}\cdot 100\%,\ (z) \in E_i. \qquad (4.21)$$

Результирующие кодовые последовательности формируют цепочку для передачи в канал связи.

6. Дробное значение шага квантователя для каждой полосы конвертируется в децибелы:

$$\Delta_{cr}(z) = 10\cdot \log_{10}\left(\delta_{CB}(z)\right),\ (z) \in E_i \qquad (4.22)$$

и кодируется шестью битами, которые также передаются в канал.

4.4. Широкополосный перцептуальный ПДВП-декодер

Структура декодера широкополосного перцептуального ПДВП-кодера речи и аудиосигналов значительно проще описанной выше структуры кодера и схематически показана на рис. 4.9.

Рис. 4.9. Структура декодера

Работа данного декодера выполняется в следующем порядке:

1. Входная информация разделяется на два битовых потока: закодированный шаг квантования и кодированная битовая последовательность отсчётов частотной полосы.

2. Шаг квантования преобразуется из децибел в разы, а уровень квантования каждого отсчета частотной полосы декодируется путем подстановки значений из кодовых книг Хаффмана соответствующей текущей битовой комбинации.

3. Восстановление вэйвлет-коэффициентов в каждой частотной полосе выполняется путем перемножения уровня квантования каждого отсчета с шагом квантования данной полосы.

4. Выполняется распределение коэффициентов по полосам согласно структуре дерева ПДВП, так как частотные полосы, в которых сигнал является неслышимым для человека – не передаются в декодер.

5. Реконструкция сигнала осуществляется синтезирующим банком цифровых фильтров, реализующим обратный ПДВП. Структура дерева обратного ПДВП соответствует структуре дерева прямого ПДВП (см. рис. 4.5), только процесс обработки сигнала происходит снизу вверх, выполняя рекурсивно базовую декомпозицию синтеза (рис. 4.10).

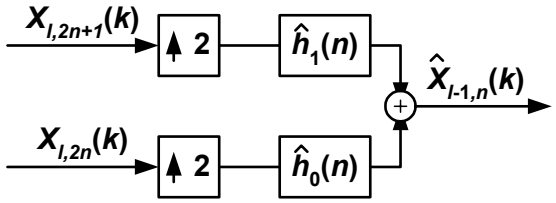


Рис. 4.10. Базовая декомпозиция синтеза ПДВП

Синтезирующая базовая декомпозиция представляется уравнением

$$X_{l-1,n}(k)=\sum_{m\in Z}h_1(m-2k)X_{l,2n}(k)+\sum_{m\in Z}h_0(m-2k)X_{l,2n+1}(k), \tag{4.23}$$

где $h_1(k)$ и $h_0(k)$ – коэффициенты высокочастотного и низкочастотного синтезирующих вэйвлет-фильтров соответственно.

При синтезе сигнала, аналогично анализу, также используется вэйвлет-функция Добеши 20-го порядка. Фреймы синтезированного сигнала домножается на взешивающую функцию окна Хеннинга и суммируются перекрывающимися частями для образования непрерывного реконструированного аудиосигнала или речи.

4.5. Расчёт порогов маскирования в вэйвлет-области

Процедура расчёта порогов маскирования в вэйвлет-области выполняется согласно приведенной ниже последовательности шагов.

1. Вычислить спектральную энергию барка (рис. 4.11):

$$A_{CB}(z)=\sum_{k=0}^{K-1}X_{z,k}^2\ , \tag{4.24}$$

где $z=\overline{1,24}$ – номер критической полосы; K – количество вэйвлет-коэффициентов $X_{z,k}$ преобразования в каждой критической полосе z (см. табл. 4.1).

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

- индекс $a_{tmn}(z)$, который оценивает отношение маскирования тоном шума, задается согласно ISO/MPEG стандарта:

$$a_{tmn}(z)=-0{,}275\cdot z-15{,}025\ \text{[дБ]},\ z=\overline{1,24}; \tag{4.25}$$

- индекс маскирования шумом шума a_{nmn} оценивается как константа

$$a_{nmn} = -25 \text{ [дБ]}; \tag{4.26}$$

• среднее значение тональности маскеров в каждой критической полосе определяется маскирующим индексом:

$$a_{CB}(z) = \eta \cdot a_{tmn}(z) + (1-\eta) \cdot a_{nmn}(z) \text{ [дБ]}, \; z = \overline{1,24}, \tag{4.27}$$

где η – тональный коэффициент:

$$\eta = \min\left(SFM_{\text{дБ}} / SFM_{\text{дБmax}}, 1\right). \tag{4.28}$$

Здесь $SFM_{дБ}$ – мера спектральной пологости; $SFM_{дБmax}$ – максимальное значение меры пологости спектра, равное -60 дБ.

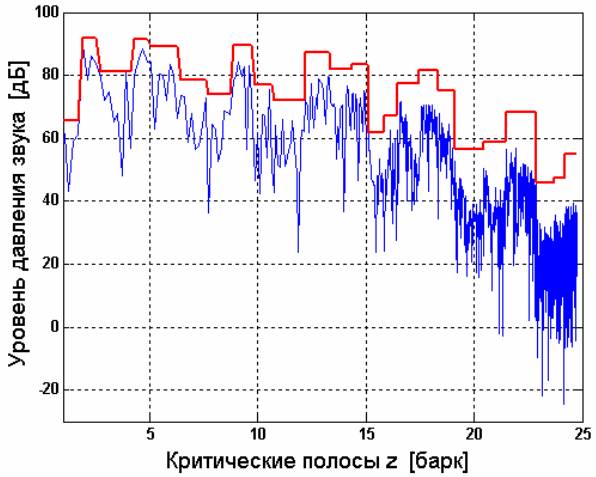


Рис. 4.11. Функции $X_{z,k}{}^2$ и $A_{CB}(z)$

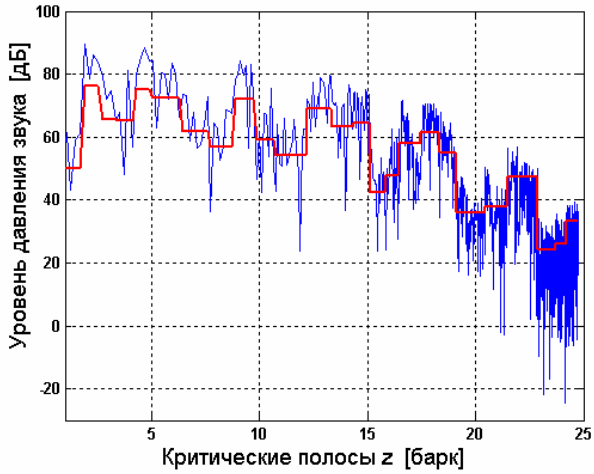


Рис. 4.12. Функции $X_{z,k}{}^2$ и $D_{CB}(z)$

3. Спектральная энергия барка с учетом тональности сигнала равна (рис. 4.12):

$$D_{CB}(z) = 10 \cdot \log\left(A_{CB}(z) \cdot 10^{\frac{a_{CB}(z)}{10}} \right) \text{ [дБ]}, \; z = \overline{1,24}. \tag{4.29}$$

4. Вычислить разброс энергии барка $C_{CB}(z)$ в каждой критической полосе z как свертку $D_{CB}(z)$ с функцией разброса $B(z)$ (рис. 4.13):

$$C_{CB}(z) = 10 \cdot \log\left(\frac{1}{K} \sum_{k=1}^{25} 10^{\frac{D_{CB}(k)}{10}} \cdot 10^{\frac{B(z-k)}{10}} \right) \text{ [дБ]}, \; z = \overline{1,24}. \tag{4.30}$$

где функция разброса $B(z)$ оценивает эффективность выполнения операции маскирования вдоль критической полосы и описывается математическим выражением

$$B(z) = a + \frac{v+u}{2} \cdot (z - z_k + c) - \frac{v-u}{2} \cdot \sqrt{\left(d + (z - z_k + c)^2\right)} \text{ [дБ]}. \tag{4.31}$$

Здесь v – нижний склон [дБ/барк]; u – верхний склон [дБ/барк]; d – пиковая плоскость; c и a – компенсирующие факторы, необходимые для сохранения соотношения $B(0) = 1$; z_k – центр маскирующего маскера.

Значения параметров для функции $B(z)$ определены в первой строке табл. 4.2.

Таблица 4.2

Функция разброса	v	u	d	c	a
Барк шкала	30 дБ/барк	-25 дБ/барк	0,3	0,05	15
Временная шкала	0,0825 дБ/F_{min}^{*}	-0,0412 дБ/F_{min}^{*}	0,3	0,157	0,032/F_{min}^{*}

F_{min}^{*} – минимальная длинна анализируемого фрейма.

5. Найти временные маскирующие пороги:

- вычислить энергию вэйвлет-коэффициента в каждой критической полосе z :

$$E_z(k) = X_{z,k}^2,\ k = \overline{0, K-1},\ z = \overline{1,25}; \tag{4.32}$$

- определить временную функцию разброса энергии в каждой критической полосе z как свертку $E_z(k)$ и функции разброса $B(k)$:

$$F_z(m) = \frac{1}{K}\sum_{k=0}^{K-1} E_z(k)\cdot 10^{\frac{B(K-k)}{10}},\ m = \overline{0, K-1}, \tag{4.33}$$

где временная функции разброса $B(k)$:

$$B(k) = a + \frac{v+u}{2}\cdot(k+c) - \frac{v-u}{2}\cdot\sqrt{\left(d+(k+c)^2\right)}\ \text{[дБ]}. \tag{4.34}$$

При этом максимальное временное разрешение для ПДВП имеет место в критических полосах верхних частот, которые имеют минимальную протяженность по времени $F_{min} = 4$ отсчета или 0,0454 мс. Следовательно, значения параметров функции разброса вдоль оси времени определяются как v = 40 дБ/мс = 0.0825 дБ/F_{min} и u = -20 дБ/мс = -0.0412 дБ/F_{min} (см. табл. 4.2, строку 2).

- определить временной фактор маскирования в полосе z как результат сравнения величин:

$$F_z(k) \geq E_z(k),\ k = \overline{0, K-1},\ z = \overline{1,24}. \tag{4.35}$$

Если данное соотношение выполняется, то в соответствующей критической полосе имеет место временное маскирование, в противном случае нет.

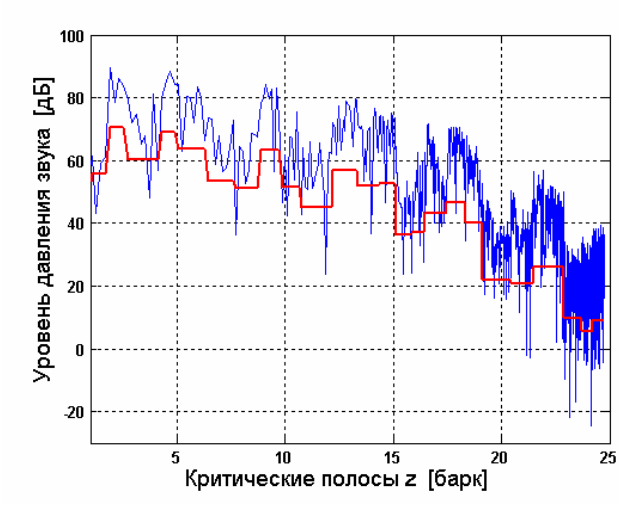


Рис. 4.13. Функции $X_{z,k}^2$ и $C_{CB}(z)$

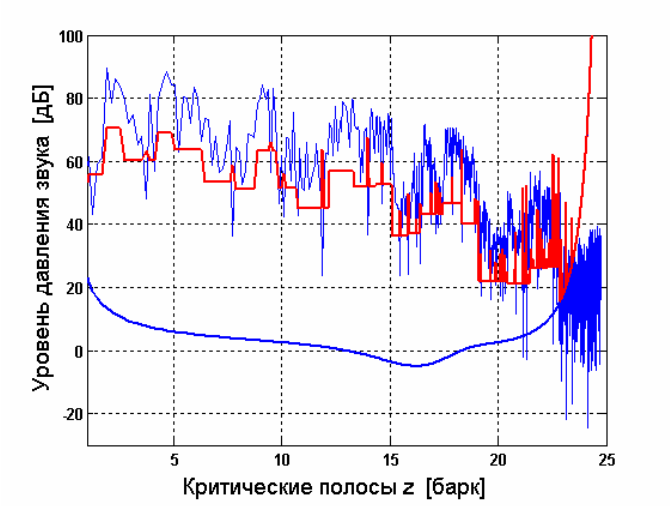


Рис. 4.14. Результирующий маскирующий порог $T_{CB}(z)$

6. Оценить частотно-временной маскирующий порог $M_{CB}(z)$ в каждой критической полосе:

$$M_{CB}(z) = C_{CB}(z) \cdot \max\left(\frac{F_z(k)}{E_z(k)}, 1\right) \text{ [дБ]}, \; k = \overline{0, K-1}, \; z = \overline{1,24}. \tag{4.36}$$

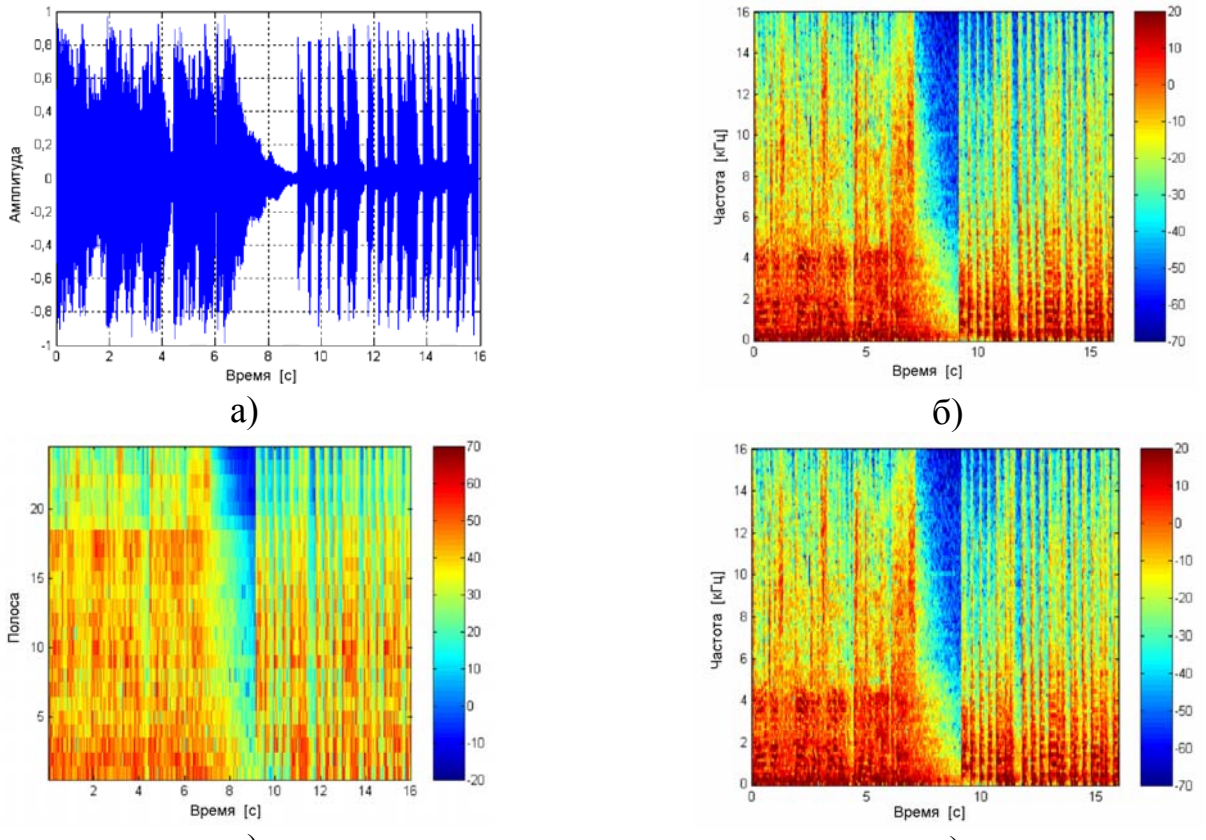
Результирующее значение маскирующего порога $T_{CB}(z)$ в соответствующей критической полосе частот получается из сравнения временно-частотного маскирующего порога $M_{CB}(z)$ с минимальным значением абсолютного порога слышимости $ATH(z)$ (рис. 4.14):

$$T_{CB}(z) = \max\left(ATH(z), M_{CB}(z)\right) \text{ [дБ]}. \tag{4.37}$$

Конец процедуры расчёта порогов маскирования в вэйвлет-области.

4.6. Экспериментальные результаты

В качестве тестового материала в экспериментах с широкополосным перцептуальным вокодером на основе ПДВП использовались как WAV-файлы музыки (16 бит, ИКМ, частота дискретизации 32 кГц),так и WAV-файлы чистой речи (16 бит, ИКМ, частота дискретизации 16 кГц). Пример обработки в данном вокодере аудиосигнала представлен на рис. 4.15.



а) б)

в) г)

Рис. 4.15. Обработка аудиосигналов в широкополосном перцептуальном кодере на основе ПДВП: а – входной аудио сигнал во временной области; б – спектрограмма входного сигнала; в – скалограмма входного сигнала на выходе анализирующего банка фильтров ПДВП и перцептуальной обработки; г – спектрограмма

восстановленного аудио сигнала (37 Кбит/с)

На рис. 4.16 показан спектр оригинального сигнала, порог маскирования и шум квантователя, который полностью замаскирован.

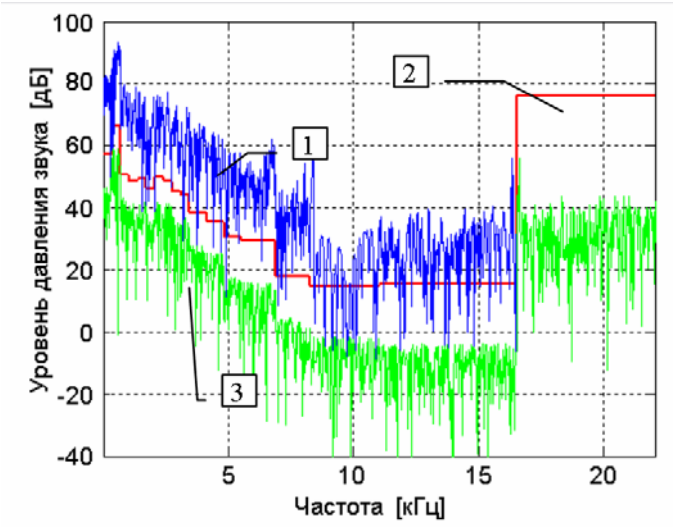
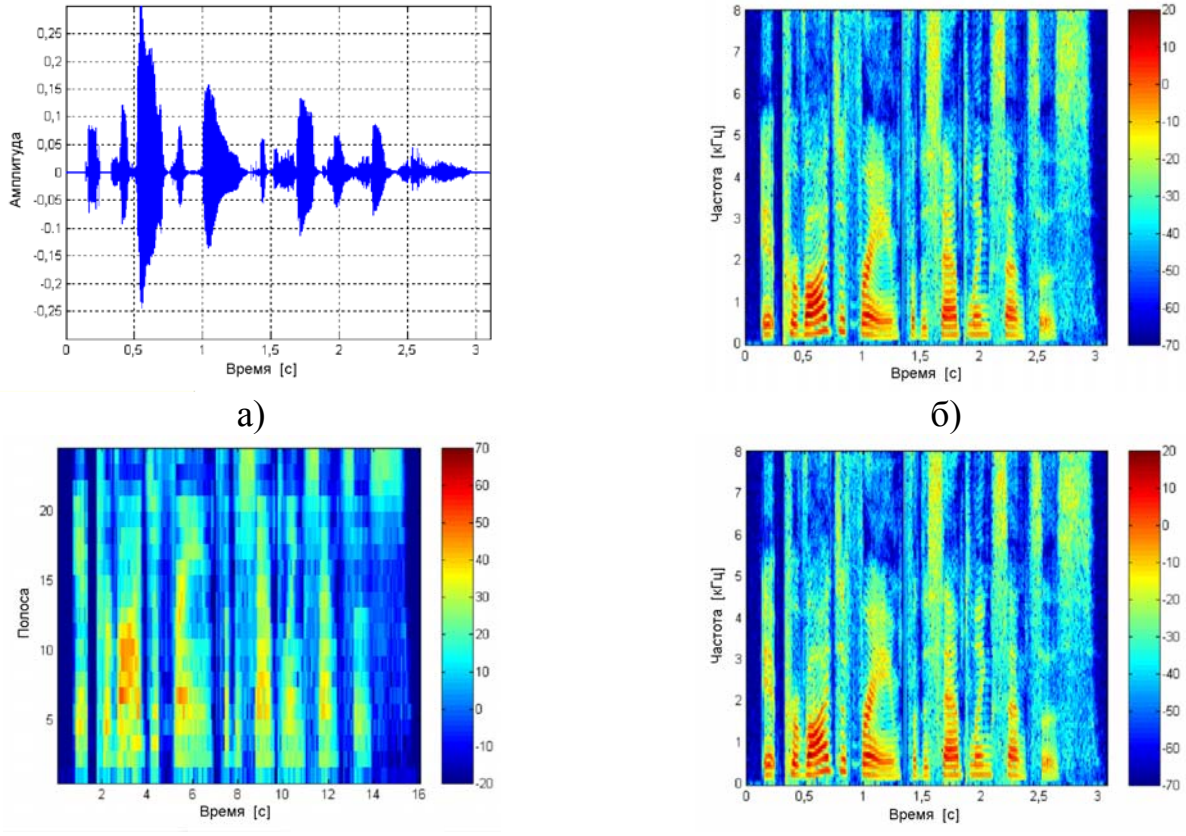


Рис. 4.16. Спектр входного аудиосигнала (1), порог маскирования (2) и шум квантователя (3)

Результаты эксперимента по компрессии речевого сигнала с помощью широкополосного перцептуального ПДВП-кодера иллюстрируется на рис. 4.17.



а) б)

в) г)

Рис. 4.17. Компрессия речевого сигнала с помощью широкополосного перцептуального ПДВП-кодера: а – входной речевой сигнал во временной области; б – спектрограмма входного сигнала; в – скалограмма входного сигнала на выходе

анализирующего банка фильтров ПДВП и перцептуальной обработки; г – спектрограмма восстановленного речевого сигнала (17 Кбит/с)

Полученные результаты позволяют судить о достаточно хорошем реконструировании входных сигналов как музыки, так и речи. Качество восстановленного сигнала для данной модели вокодера практически неразличимо с оригиналами аудио- и речевого сигналов. При этом степень компресии достигается порядка 10-15 раз, что составляет 34-50 Кбит/с для аудиосигналов и 17-25 Кбит/с для сигналов речи. Дополнительный выигрыш в компрессии речевого сигнала может быть достигнут за счет применения схем детектирования пауз (интервалов отсутствия речи)

ЗАКЛЮЧЕНИЕ

В настоящем методическом пособии по курсу «Речевые интерфейсы ЭВС» проведен анализ существующих методов кодирования речевых данных, а также описаны особенности речевых сигналов и их обработки. Большое внимание уделено механизмам восприятия речи человеком. Рассмотрены базовые концепции современной психоакустики и численные модели, используемые в ней. Показано, что учет психоакустических закономерностей позволяет при неизменной скорости передачи обеспечить лучшее перцептуальное качество синтезированной речи.

Рассмотрено алгоритмическое обеспечение и программная модель широкополосных вокодеров на основе CELP-модели с многополосным возбуждением и перцептуальной оптимизацией с диапазоном частот 50...7500 Гц. Данные вокодеры имеют гибкую структуру, позволяющую менять параметры кодирования для получения необходимого качества реконструированной речи при заданной пропускной способности коммуникационного канала. Описан перцептуальный широкополосный кодер речевых сигналов на основе пакета дискретного вэйвлет-преобразования. Приводится процедура расчета психоакустической модели в вэйвлет-области.

Планируется выпуск серии пособий по речевым интерфейсам ЭВС, в которых, в частности, будут затронуты вопросы векторного квантования, модели речеобразования, основанные на схеме согласованной с основным тоном декомпозиции речевого сигнала на периодический и шумовой компоненты и ее использование в компрессии речи и синтезе речи по тексту, а также перцептуальное аудиокодирование.

ЛИТЕРАТУРА

1. Рабинер Л.Р., Шафер Р.В. Цифровая обработка речевых сигналов. М.: Радио и связь, 1981. – 495 с.
2. Маркел Дж., Грэй А. Линейное предсказание речи. М.: Связь, 1980.
3. Марпл-мл. С. Цифровой спектральный анализ и его приложения. М.: Мир, 1990.
4. Назаров М.В., Прохоров Ю.Н. Методы цифровой обработки и передачи речевых сигналов. М.: Радио и связь, 1985.
5. Прохоров Ю.Н. Статистические модели и рекуррентное предсказание речевых сигналов. М.: Радио и связь, 1984.
6. Сапожков М.А., Михайлов В.Г. Вокодерная связь. М.: Радио и связь, 1983.
7. Шелухин О.И., Лукьянцев Н.Ф. Цифровая обработка и передача речи. М.: Радио и связь, 2000.
8. Kondoz A.M. Digital Speech Coding for Low Bit Rate Communications Systems. UK University of Surrey: John Wiley & Sons, 1996. –442 p.
9. Speech coding and synthesis // Edited by W.B.Kleijn, K.K.Paliwal. Amsterdam: Elsevieer, 1998.
10. Barnwell T.P., Nayebi K., Richardson C.H. Speech coding: a computer laboratory textbook. NJ: J.Wiley & Sons, 1996.
11. Hess W.J. Pitch and voicing determination, in Advances in speech signal processing, Sadaoka Furui and M. M. Sohndi, Eds. New York: Marcel Dekker, 1992. P. 3-48.
12. Mobile Radio Communications // By Raymond Steele edition, PENTECH PRESS Publishers – London, 1992.
13. Atal B.S., Schroeder M.R. Predictive coding of speech and subjective error criteria // IEEE Trans. ASSP. Vol. 27. № 3. 1979. June. – P. 247-254.
14. Atal B.S., Schroeder M.R. Optimizing predictive coders for minimum audible noise // Proc. ICASSP'79. 1979. – P. 453-455.
15. Применение цифровой обработки сигналов // Под ред. Э.Оппенгейма. М.: Мир. 1980. С. 60-66; 137-191.
16. Zwicker E., Fastl H. Psychoacoustics: Facts and Models. Springer-Verlag Berlin Heidelberg, 1990.
17. Johnston J.D. Transform coding of audio signals using perceptuals noise criteria // IEEE Trans. on Select. Areas Commun. 1988. Feb. Vol. 6. P. 314-323.
18. Daubechies I. Ten lectures on Wavelets. // Society for industrial and applied mathematics. Philadelphia, Pennsylvania, 1992. – 357 p.
19. Petrovsky Al., Krahe D., Petrovsky A.A. Real-Time Wavelet Packet-based Low Bit Rate Audio Coding on a Dynamic Reconfigurable System // Proc. of the 114th AES Convention. Preprint № 5778. 22-25 May, Amsterdam, Netherlands, 2003. – 22 p.
20. ITU-T Recommendations on CD-ROM, 1998 June. // International Telecommunications Union, the Telecommunication Standardizations Sector of the International Telephone and Telegraph Consultative Committee, CCITT.

Приложение 1. АНАЛИЗИРУЮЩИЙ-СИНТЕЗИРУЮЩИЙ ДПФ-БАНК ПОЛИФАЗНЫХ ФИЛЬТРОВ

Пусть порядок КИХ-фильтра N кратен коэффициенту прореживания q. Одномерную последовательность коэффициентов фильтра $h_n=h(n)$, $n=0,...,N-1$ (каждому отсчету импульсной характеристики $h(n)$ ставится в соответствие коэффициент h_n)), представим в виде двухмерной матрицы коэффициентов размерности q x L: $[h_{k,l}]$, где $h_{k,l}=h(k+ql)$; $l=0,...,L-1$; $k=0,...,q-1$. Нулевая строка $k=0$ матрицы $[h_{k,l}]$ получается простым прореживанием последовательности $h(n)$, $n=0,...,N-1$ с коэффициентом прореживания q, а каждая последующая строка предполагает сдвиг влево последовательности $h(n)$ на число отсчетов, определяемый номером строки.

Таким образом, каждой k-ой строке матрицы коэффициентов $[h_{k,l}]$ можно поставить в соответствие некоторый КИХ фильтр L-го порядка с передаточной функцией $H_{k,l}(z^{-1})$ и импульсной характеристикой $h_k(l)$, $l=0,...,L-1$, отличающимися тем, что частота дискретизации как входной, так и выходной последовательности данных равна f_s/q, где f_s - частота дискретизации входной последовательности $x(n)$. Окончательный результат вычисления, совпадающий с реакцией $y(nq)$ фильтра дециматора N-го порядка в моменты времени $n = qm$, формируется суммированием выходных последовательностей $y_k(m)$ всех q полифазных фильтров.

Для M-канального банка ДПФ-фильтров частотная характеристика k-го фильтра нижних частот формируется путем сдвига на величину $2\pi k/M$:

$$H_{k,l} = H_0\left(z \cdot e^{-j\frac{2\pi k}{M}}\right), \qquad (П1.1)$$

а импульсная характеристика определяется выражением

$$h_{k,l}(n) = h_0 \cdot e^{j\frac{2\pi k}{M}}. \qquad (П1.2)$$

На рис. П1.1 показана полифазная форма банка ДПФ-фильтров из одного фильтра.

Таким образом, фильтр нижних частот с коэффициентами $h_0(n)$ разделяется на M-фаз:

$$H_0(z) = E_0(z^M) + z^{-1}E_1(z^M) + ... + z^{-1}E_{M-1}(z^M). \qquad (П1.3)$$

Матрица анализирующего банка ДПФ-фильтров на основе полифазной структуры определяется следующим образом:

$$H_k(z) = [DFT]diag(E_0(z), E_1(z), ..., E_{M-1}(z)). \qquad (П1.4)$$

Идеализированная частотная характеристика банка ДПФ фильтров показана на рис. П1.2.

Построение синтезирующего банка ДПФ-фильтров на основе полифазной структуры осуществляется в обратном порядке конструирования анализирующего

банка. Передаточная функция синтезирующего ДПФ-банка полифазных фильтров есть:

$$F_0(z) = R_{M-1}(z^M) + z^{-1}R_{M-2}(z^M) + \ldots + z^{-1}R_0(z^M). \quad (П1.5)$$

Результат анализа и синтеза есть диагональная матрица, когда выполнится прямое и обратное ДПФ:

$$F_k(z)H_k(z) = diag(R_0(z)E_0(z), R_1(z)E_1(z), \ldots, R_{M-1}(z)E_{M-1}(z)). \quad (П1.6)$$

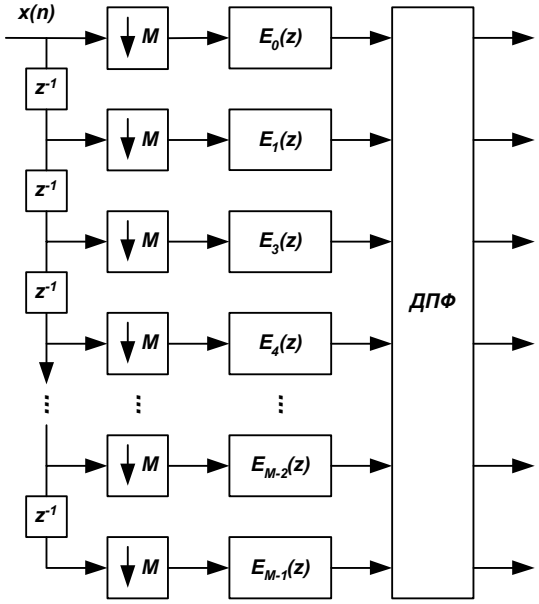


Рис. П1.1. Полифазная структура банка ДПФ-фильтров

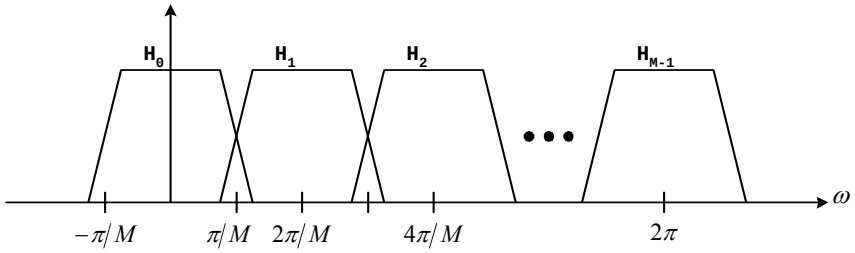


Рис. П1.2. Идеализированная частотная характеристика банка ДПФ-фильтров.

Структура *M*-канального анализирующего/синтезирующего ДПФ-банка полифазных фильтров приведена на рис. П1.3.

Представленное выше преобразование полифазной формы построения *M*-канальной системы было рассмотрено для модели комплексного входного сигнала с равномерным расположением *M* субполос в диапазоне частот $0 \le \omega \le 2\pi$. Для модели действительного входного сигнала с равномерным распределением *M* субполос в диапазоне частот $0 \le \omega \le \pi$ коэффициент прореживания отсчетов выходных сигналов $q=2M$ и число полифазных фильтров равно *2M*. Разделение частотных каналов выполняется с использованием *2M*-точечного ДПФ-преобразования по *M*-выходам.

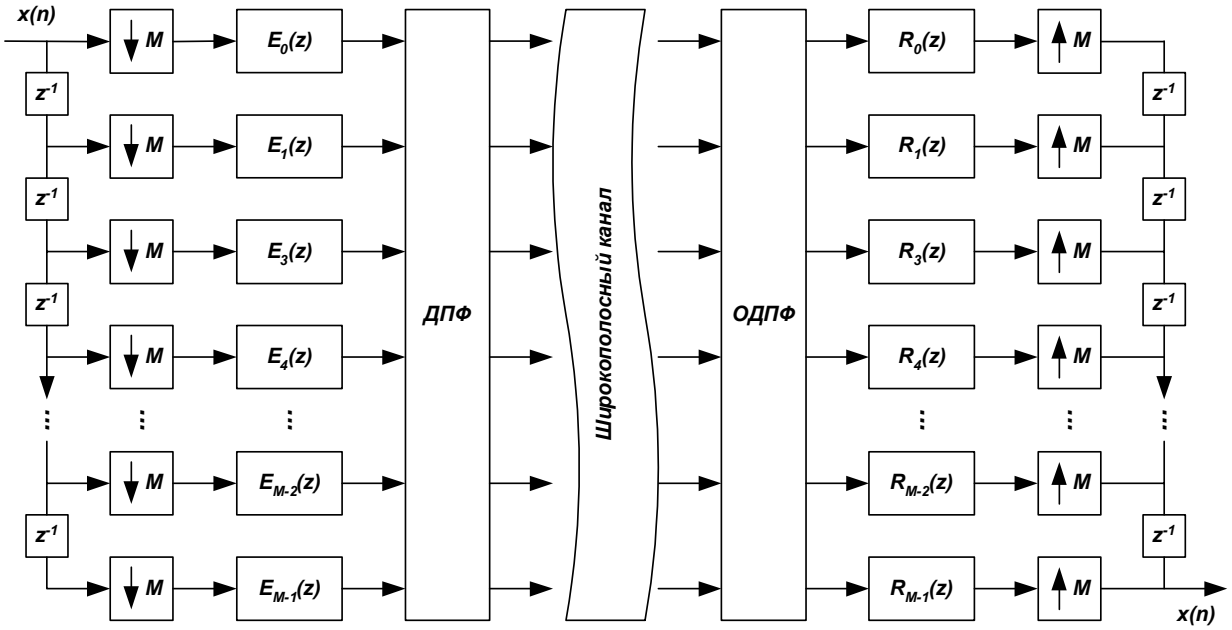


Рис. П1.3 Структура М-канального анализирующего/синтезирующего ДПФ-банка полифазных фильтров

Приложение 2. ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЕ И ПАКЕТ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЯ

Новое теоретическое направление цифровой обработки сигналов – дискретное вэйвлет-преобразование, которое позволяет осуществлять одновременно обработку нестационарных сигналов в частотной и временной областях, проводить анализ сигналов с переменным время частотным разрешением. Дискретное вэйвлет-преобразование описывается следующими итерационными базовыми уравнениями:

- масштабная функция:

$$\varphi(t) = \sqrt{2}\sum_n h_0(n)\varphi(2t-k)\,, \qquad (\text{П2.1})$$

где $h_0(n)$ - последовательность реальных или комплексных чисел, называемых коэффициентами фильтра масштабной функции; $\sqrt{2}$ - нормирующий коэффициент масштабной функции;

- вейвлет функция:

$$\psi(t) = \sqrt{2}\sum_n h_1(n)\varphi(2t-k)\,, \qquad (\text{П2.2})$$

где $h_1(n)$ - коэффициенты фильтра вейвлет-функции.

После выполнения масштабирования, сдвига значений по времени и замены переменной $m=2k+n$ в (П2.1) получим следующее выражение для масштабной функции:

$$\varphi(2^j t-k) = \sqrt{2}\sum_m h_0(m-2k)\varphi(2^{j+1}t-m)\,. \qquad (\text{П2.3})$$

Если определить пространство V_j как

$$V_j = \underset{k}{Span}\left\{2^{j/2}\varphi\left(2^j t - k\right)\right\}, \tag{П2.4}$$

то все V_j образуют вложенные ограниченные пространства

$$... \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_{+1} \subset V_{+2} \subset ... \subset L^2 \tag{П2.5}$$

или

$$V_j \subset V_{j+1} \text{ для всех } j \in Z, \tag{П2.6}$$

т.е. пространство с высоким уровнем разрешения включает пространство с низким уровнем разрешения. Согласно (П2.6) должно удовлетворяться условие

$$x(t) \in V_j \Leftrightarrow x(2t) \in V_{j+1}, \tag{П2.7}$$

тогда входной сигнал *x(t)* может быть представлен как

$$x(t) = \sum_k c_{j+1}(k) 2^{(j+1)/2} \varphi\left(2^{j+1} t - k\right), \tag{П2.8}$$

где c_{j+1} – масштабные коэффициенты уровня $j+1$, которые вычисляются на основе масштабной функции.

Для детального представления сигнала $x(t)$ на уровне j вместе с масштабной функцией необходимо применение вейвлетной функции, которая вычисляется из (П2.2) тем же методом, что и масштабная. Следовательно, для уровня j сигнал представляется следующим соотношением:

$$x(t) = \sum_k c_j(k) 2^{j/2} \varphi\left(2^j t - k\right) + \sum_k d_j(k) 2^{j/2} \psi\left(2^j t - k\right), \tag{П2.9}$$

где $2^{j/2}$ - нормирующий коэффициент.

Масштабные c_j и вейвлетные d_j коэффициенты уровня j определяются из результирующих коэффициентов масштабной и вейвлетной функции уровня $j+1$:

$$c_j = \sum_m h(m - 2k) c_{j+1}(m) \text{ и } d_j = \sum_m g(m - 2k) d_{j+1}(m). \tag{П2.10}$$

Из (П2.10) следует, что коэффициенты уровня j фильтруются двумя фильтрами низких и верхних частот с конечной импульсной характеристикой, заданными коэффициентами $h(n)$ и $g(n)$. Далее, выполняя операции децимации обеих частей преобразования, получают коэффициенты масштабной и вейвлетной функций для следующей ступени дерева дискретного вейвлет-преобразования.

Характеристики вейвлет-преобразования рассмотрим на основе ортогонального вейвлет представления. Предположим, что вейвлет-функция $\psi(x)$ компактно представлена в интервале $[a, b]$, тогда $\psi_{s,u}(x)$ представима в $[(a+u)2^s,(b+u)2^s)]$. Допустим также, что ваниш-моменты вэйвлет-функции $\psi(x)$ располагаются в границах Фурье области $[\omega_{\min}, \omega_{\max}]$. Возможность частотной локализации является важной характеристикой вейвлет-представления. Предположим, что амплитуда на частоте ω_0 вызывает интерес, тогда вейвлетные коэффициенты масштаба s

$$\log_2(\omega_{\min}/\omega_0) \le s \le \log_2(\omega_{\max}/\omega_0) \qquad (П2.11)$$

будут использованы для вычисления интересующего отсчета на ω_0. Так как частотная декомпозиция, определяемая вейвлет-преобразованием, является фиксированной, то пакет дискретного вэйвлет-преобразования (ПДВП) может быть применим для исследования сигналов в частотной области. На базе вейвлет-преобразования строится ПДВП, который описывается выполнением декомпозиции всех ветвей дерева вэйвлет-преобразования, как показано на рис. П2.1.

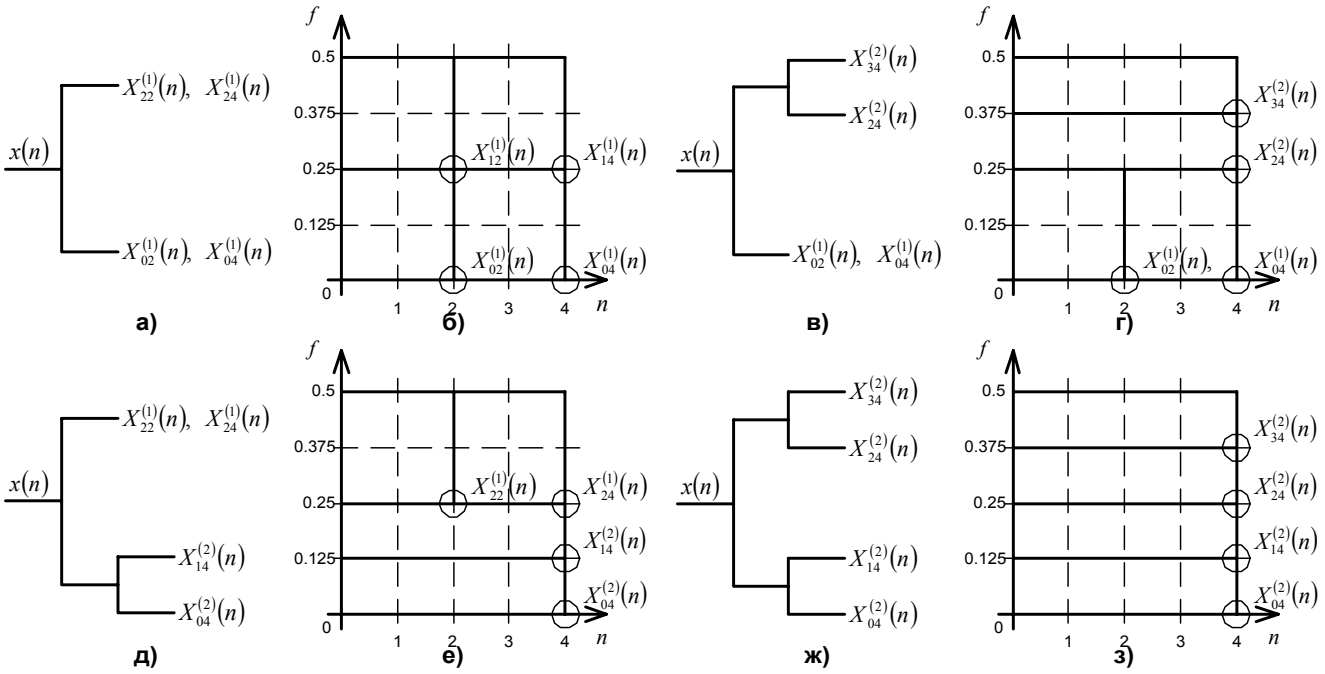


Рис. П2.1. Пример структур деревьев ПДВП для уровня 2 и их карты частотно-временной локализации

Приложение 3. ОБЗОР МЕТОДОВ ПОСТРОЕНИЯ ВОКОДЕРОВ

Все методы кодирования можно разделить на два класса: кодеры формы сигнала и параметрические кодеры. В кодерах формы сигнала ставится задача максимально точного воспроизведения речи во временной области. Эти кодеры и схемы, основанные на линейном предсказании (LPC-10, CELP, MELPC, RPELPC), обеспечивают приемлемое качество при скоростях передачи 6 Кбит/с и выше. В диапазоне скоростей ниже 4 Кбит/с, который требуется для беспроводных и спутниковых систем коммуникаций, при сохранении разборчивости у данных методов отмечается существенная деградация качества синтезируемой речи.

Параметрическими называют кодеры, у которых синтезируемый речевой сигнал при уменьшении ошибок квантования не сходится к оригинальному. При параметрическом кодировании каждый фрейм речевого сигнала характеризуется набором параметров модели речеобразования, квантуемых без учета оригинального сигнала. Качество синтезируемой речи для таких кодеров ограничивается точностью модели. Поскольку восстанавливаемый сигнал даже при хорошем субъективном качестве сильно отличается от входного, соотношение сигнал–шум не может

использоваться как характеристика кодера. Представляется естественным, что модель параметрических кодеров должна быть основана на физиологической структуре речеобразующего аппарата человека. При этом становится возможной эмуляция характеристик голосового тракта и звуковой волны, образуемой голосовыми связками. Однако в большинстве практических случаев от моделирования физиологии отказываются ради обеспечения меньшей вычислительной сложности кодера и повышения качества синтезируемой речи.

Многочисленные исследования показали, что при низких скоростях передачи более эффективным является параметрическое кодирование, при котором модель речеобразования учитывает компактный набор параметров, особенно значимых для восприятия. Наибольшее распространение получили кодеры, использующие в качестве параметров различные характеристики представления сигнала в частотной области – спектральные кодеры. Исследования сосредоточились в основном на группе гибридных кодеров, родоначальниками которой считаются Harmonic Coding (HC), Sinusoidal Transform Coding (STC) и Multiband Excitation Coding (MBE).

LPC–вокодеры. LPC-анализ осуществляется на фрейме речевого сигнала, информация о параметрах предсказания квантуется и передается. Осуществляется определение: вокализованный или невокализованный фрейм. Решение может быть основано или на оригинальной речи, или на остаточном сигнале предсказателя, но всегда – на степени периодичности сигнала. Если фрейм классифицируется как невокализованный, то сигнал возбуждения – белый шум. Если фрейм классифицирован как вокализованный, то передается период основного тона и сигнал возбуждения представляет собой периодический набор импульсов. В любом случае амплитуда выходного сигнала выбирается таким образом, чтобы его энергия совпадала с энергией сигнала оригинальной речи. Структурная схема данного вокодера представлена на рис. П3.1.

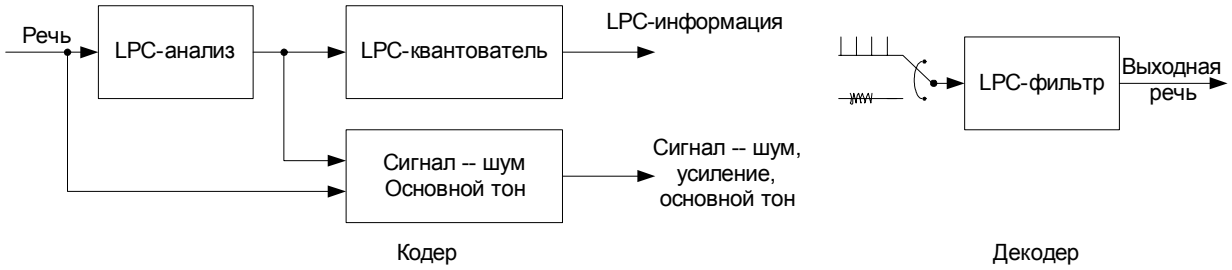


Рис. П3.1. Структурная схема LPC-вокодера

Вокодеры с многополосным возбуждением (MBE). Основой таких вокодеров является то, что речевой сигнал может быть смоделирован как комбинация гармонически связанных синусоидальных сигналов и узкополосного шума. Внутри заданной полосы речь классифицируется как периодическая или апериодическая. Гармонически связанные синусоиды используются для генерации периодических компонентов, а белый шум – для генерации апериодических составляющих. Таким образом, вместо передачи одного решения вокализованный/невокализованный фрейм состоит из набора аналогичных решений, соответствующих разным полосам. Кроме того, приемнику передаются

спектральная огибающая и коэффициент усиления. Линейное предсказание может использоваться, а может и не использоваться для квантования спектральной огибающей. Наиболее часто анализ осуществляется при помощи быстрого преобразования Фурье (БПФ). Синтез в декодере обычно осуществляется при помощи нескольких параллельных синусоидальных генераторов и генераторов белого шума. MBE–вокодеры не передают фазу синусоидальных составляющих и не пытаются выделить ничего, кроме энергии апериодических компонент. Структура вокодера с мультиполосным синусоидальным возбуждением представлена на рис. П3.2.

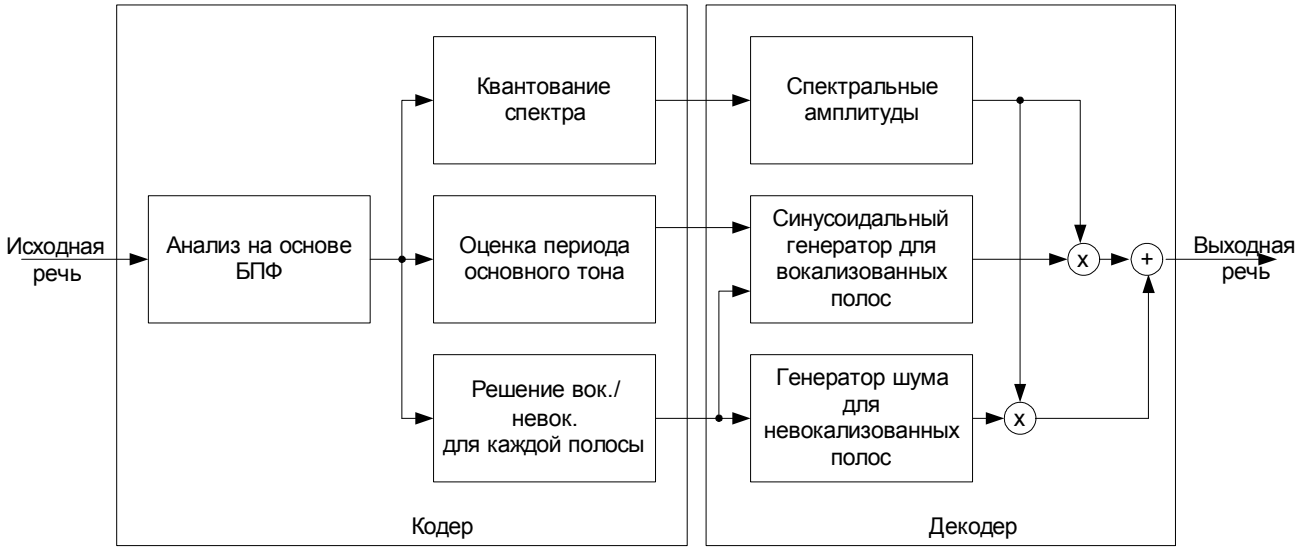


Рис. П3.2. Структура вокодера с многополосным возбуждением

Кодеры на основе интерполяции огибающей волны. В данном кодере предполагается, что речь состоит из медленно развивающейся периодической огибающей (МРО) (от англ. «slowly evolving periodic waveform – SEW») и быстроразвивающейся шумоподобной огибающей (БРО) (от англ. «rapidly evolving waveform – REW»). Фрейм анализируется с целью извлечения «характеристической огибающей», затем полученная информация фильтруется для выделения БРО из МРО. При этом информация о БРО обновляется более часто, чем о МРО. Коэффициенты линейного предсказания, частота основного тона, спектр МРО и БРО, а также общая энергия передаются независимо. В приемнике осуществляется воссоздание параметрического представления о МРО и БРО, суммирование и фильтрация с помощью синтезирующего фильтра для получения выходного речевого сигнала. Структурная схема кодера данного типа представлена на рис. П3.3, а структура декодера – на рис. П3.4.

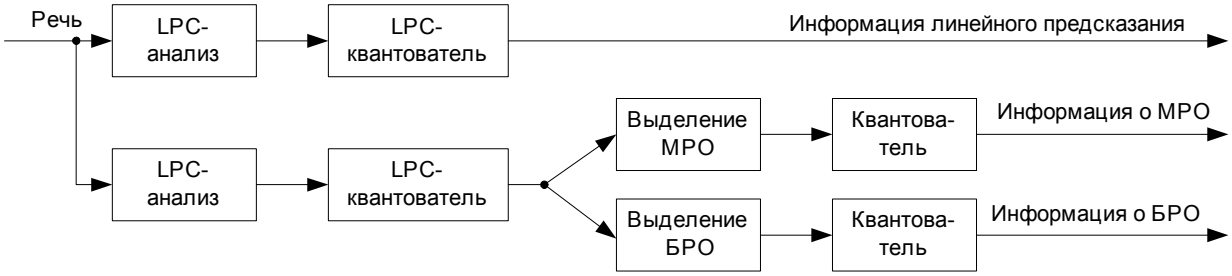


Рис. П3.3. Структурная схема кодера на основе интерполяции огибающей

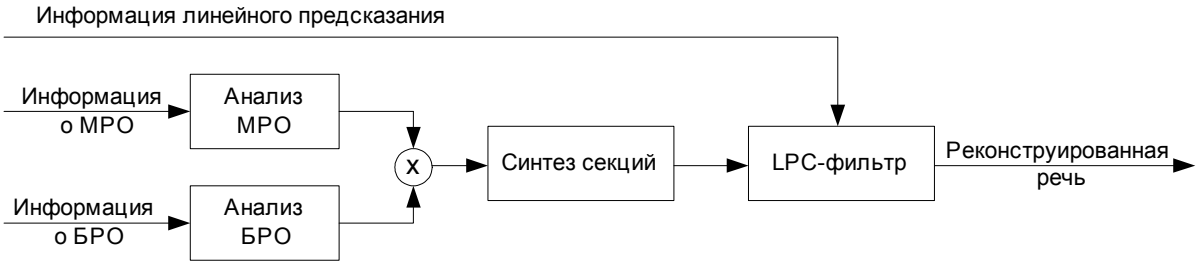


Рис. П3.4. Структурная схема декодера на основе интерполяции огибающей

Речевые кодеры с линейным предсказанием на основе схемы «анализ через синтез» (LPAbSSC). Процесс кодирования начинается с анализа на основе линейного предсказания. Типичным для данных кодеров является то, что информация о линейном предсказании определяется по методу адаптации по будущему, но существуют и исключения. LPAbSSC заимствуют концепцию наличия декодера в самом кодере от АДИКМ. Разность между выходным заквантованным сигналом и оригинальным сигналом взвешивается при помощи перцептуального фильтра. Таким образом, подбирается возбуждающий сигнал, обеспечивающий минимальное среднеквадратическое отклонение (СКО) от оригинала. Долгосрочный фильтр удаляет долговременную корреляцию, обусловленную периодичностью основного тона, из сигнала. Если в кодере присутствует структура для определения основного тона, то параметры фильтра долгосрочного предсказателя вычисляются первыми. Наиболее общей используемой системой является адаптивная кодовая книга, в которой сохраняются отсчеты предыдущей возбуждающей последовательности. В кодере выбираются, квантуются и передаются период основного тона и коэффициент усиления, которые соответствуют наименьшей перцептуальной ошибке. Следующим видом является возбуждение на основе фиксированной кодовой книги, при этом вектор возбуждения выбирается аналогично и передаются его индекс и коэффициент усиления Обобщенная структурная схема вокодеров данного типа представлена на рис. П3.5.

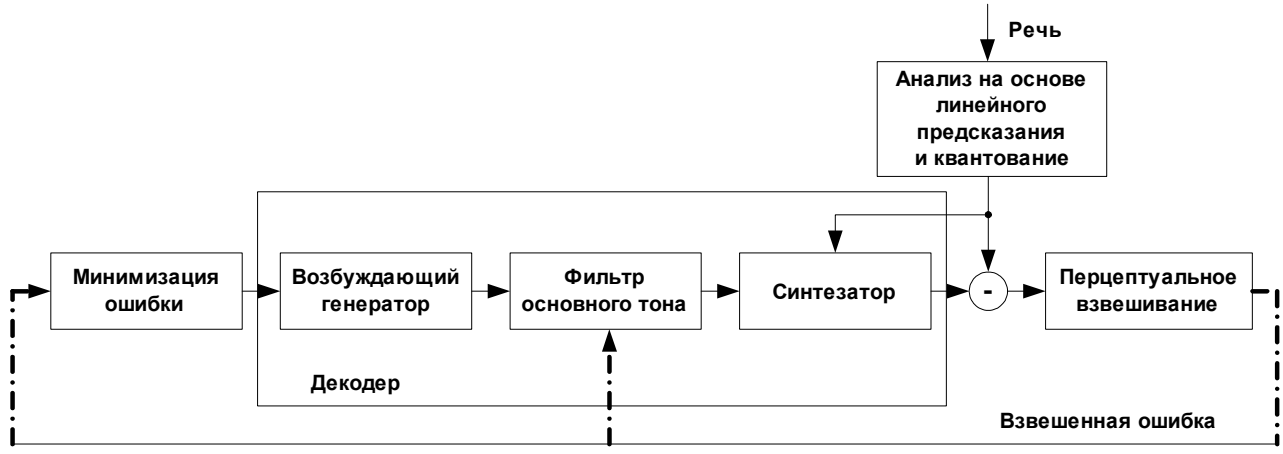


Рис. П3.5. Обобщенная структурная схема вокодеров с линейным предсказанием на основе схемы «анализ через синтез»

Множество различных фиксированных кодовых книг с соответствующими названиями было создано для кодеров, относящихся к данному классу. Коснемся лишь основных.

Кодирование на основе линейного предсказания с многоимпульсным возбуждением (MPLPC). Данный метод подразумевает, что фрейм исходного сигнала делится на более короткие субфреймы. Фиксированная кодовая книга состоит из набора импульсов, оставшихся после определения вклада адаптивной кодовой книги. Часто количество импульсов выбирается равным 0,1 от количества отсчетов в субфрейме. При кодировании сначала выбирается импульс, который вносит наибольший вклад в уменьшение ошибки, затем – импульс, вносящий следующий наибольший вклад и т.д. Как только набрано необходимое количество импульсов, определение последовательности завершено. Для каждого импульса передается его расположение в последовательности и амплитуда.

Кодирование на основе линейного предсказания с возбуждением по кодовой книге (CELP). Кодеры данного типа имеют фиксированную кодовую книгу, которая состоит из векторов. В первых CELP-кодерах кодовые книги были составлены из гауссовых случайных величин. Впоследствии было выявлено, что кодовые книги, состоящие из центрированных случайных величин, обеспечивают лучшее качество реконструированной речи. Это имело место при создании кодовых книг, похожих на набор возбуждающих векторов для мультиимпульсного возбуждения. Единственный способ увеличения скорости поиска в таких кодовых книгах – это составление кодовых книг из перекрывающихся векторов.

Кодирование на основе линейного предсказания с возбуждением векторной суммой (VSELP). Фиксированная кодовая книга состоит из взвешенной суммы набора базовых векторов. Базовые векторы являются ортогональными по отношению друг к другу. Вес любого базового вектора может быть равен либо минус 1, либо плюс 1. Быстрый поиск в таких книгах может обеспечиваться при использовании метода исследования на основе псевдокода Грея. VSELP был использован в некоторых стандартах сотовых телефонов первого и второго поколений.

Субполосные кодеры (SBC). Концепция субполосного кодирования очень проста и заключается в разделении речевого сигнала на несколько частотных полос с раздельным квантованием последних. При этом шум квантования остается внутри каждой полосы. Для разделения на полосы используются зеркальные квадратурные фильтры или банки фильтров на основе вейвлет–преобразования. Это дает несколько преимуществ: ошибка квантования всех наложений в синтезирующем банке фильтров ликвидируется вследствие децимации в анализирующем банке; полосы могут быть оцифрованы с критической частотой, т.е. количество отсчетов в частотной области будет соответствовать количеству отсчетов во временной области.

Эффективность таких кодеров в большой степени зависит от вида алгоритма квантования. В общем, алгоритмы, которые динамически размещают биты в соответствии с текущей спектральной характеристикой речи, обеспечивают

наилучшую производительность. Структурная схема субполосного вокодера представлена на рис. П3.6.

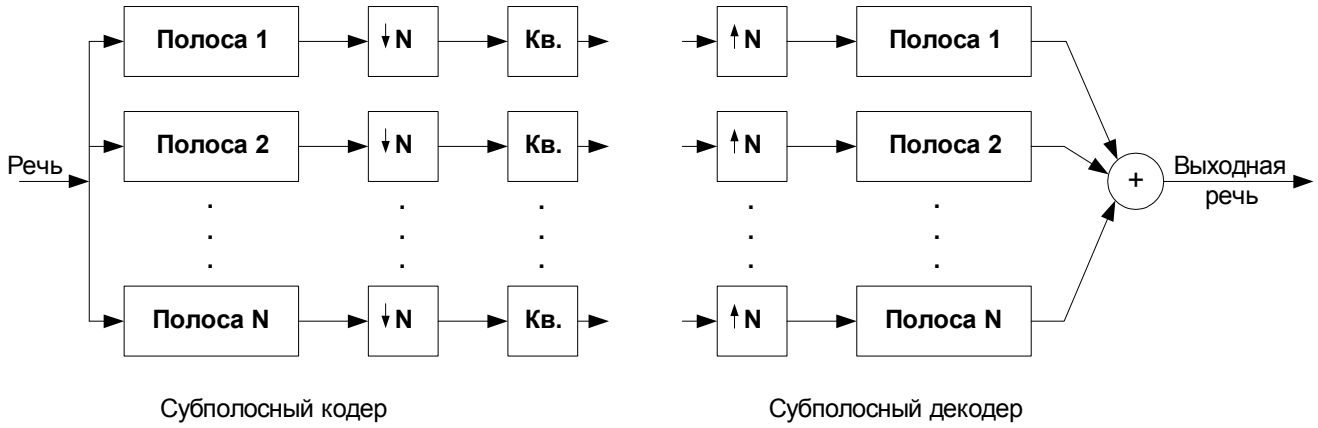


Рис. П3.6. Структурная схема субполосного вокодера

Качество синтезированного речевого сигнала обычно оценивается путем усреднения оценок специалистов по тесту MOS (Mean Opinion Score, метод мнений) в диапазоне от 0 до 5. Во многих случаях бывает удобно проводить не обобщенную оценку качества, а различать такие характеристики, как разборчивость и натуральность речи. Например, для военных приложений большую роль играют разборчивость и скорость передачи, возможно в ущерб естественности звучания. На рис. П3.7 представлены оценки натуральности и разборчивости для типичных представителей различных групп методов кодирования.

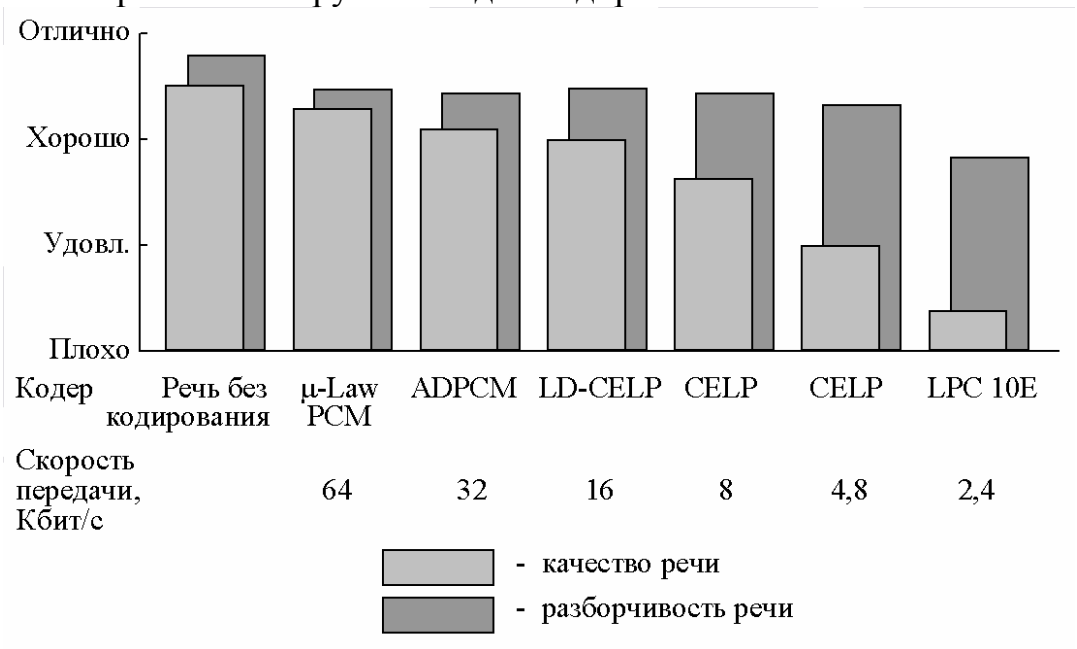


Рис. П3.7. Оценки качества различных схем кодирования

Приложение 4. ДЕЙСТВУЮЩИЕ СТАНДАРТЫ

Все ныне действующие стандарты в области телекоммуникаций регламентируются Международным союзом электросвязи (МСЭ, ITU), контролирующим все аспекты стандартизации телекоммуникационных и

радиосетей. Рекомендации МСЭ для кодеров на основе линейного предсказания, выполненных по схеме «анализ через синтез», приведены в табл. П4.1.

В табл. П4.2 приведены краткие характеристики речевых кодеров первого и второго поколений, стандартизованных для цифровой сотовой телефонии. Кодеры первого поколения обеспечивают адекватное качество. Два кодера второго поколения, так называемые полускоростные кодеры, были разработаны с целью удвоения емкости канала. Следующее поколение кодеров, появившееся вскоре после полускоростных кодеров, было предназначено для улучшения качества речи цифрового сотового сервиса до качества проводных телефонных сетей.

Ниже приводится краткая информация о стандартах безопасной (секретной) голосовой связи. **FS1015** – Федеральный стандарт США для LPC-вокодера со скоростью 2,4 Кбит/с. Данный стандарт предназначен для терминалов, использующих секретную связь. Вокодеры данного типа не обеспечивают естественного звучания речи, но в процессе эволюции качество существенно повысилось. Наиболее распространенная версия вокодера по стандарту FS1015 со скоростью 800 бит/с с применением векторного квантования была стандартизована НАТО.

Таблица П4.1

Рекомендации МСЭ для кодеров на основе линейного предсказания, выполненных по схеме «анализ через синтез»

Стандарт	ITU G.728	ITU G.729	ITU G.723.1
Год	1992 и 1994	1995	1995
Тип кодера	LD-CELP	CS-ACELP	MPC-MLQ и ACELP
Скорость, Кбит/с	16	8	6,3 и 5,3
Качество	С потерями	С потерями	С небольшими потерями
Сложность: MIPS / RAM	30 / 2 К	<22 / 2.5 К	<16 / 2.2 К
Задержка: размер фрейма / упреждающий буфер	0.625 мс / 0 мс	10 мс / 5 мс	30 мс / 7,5 мс
Тип спецификации: плавающая запятая / фиксированная запятая	Алг. точность / до бита	Нет / до бита	Нет / до бита

Федеральный стандарт **FS1016** является результатом проекта, порученного Департаменту защиты для повышения естественности речи в секретном телефонном блоке (STU-3) с внедрением модемной технологии для скорости 4,8 Кбит/с. В качестве базового был выбран кодер на основе CELP-алгоритма, имеющий так называемую тройную книгу, в которой все амплитуды возбуждающего сигнала равны минус 1, 0 или плюс 1 перед масштабированием на коэффициент усиления для обрабатываемого субфрейма. FS1016, несомненно, сохраняет большую реалистичность оригинальной речи, чем FS1015, но речь все еще содержит много артефактов и качество её в основном ниже, чем у сотовых кодеров стандарта IS-54.

Следующий кодер, который был стандартизован Департаментом защиты, - это кодер со скоростью 2,4 Кбит/с, заменяющий FS1015 и FS1016. В качестве основы выбран MELP-кодер. Необходимость в таком кодере возникла из-за отсутствия достаточного количества спутниковых каналов, обеспечивающих пропускную

способность 4,8 Кбит/с. Качество данного кодера аналогично FS1016, а иногда и превышает его.

Таблица П4.2

Речевые кодеры цифровой сотовой телефонии

Стандартизатор	CEPT	ETSI	TIA	TIA	RCR	RCR
Название стандарта	GSM	GSM полу-скоростной	IS-54	IS-96	PDC	PDC полу-скоростной
Тип кодера	RPE-LTP	VSELP	VSELP	CELP	VSELP	PSI-CELP
Год	1987	1994	1989	1993	1990	1993
Скорость [Кбит/с]	13	5,6	7,95	0,8...8,5	6,7	3,45
Качество	С потерями	GSM	GSM	Хуже GSM	Хуже GSM	Хуже GSM
Сложность: MIPS / RAM	4,5 / 1 K	30 / 4 K	20 / 2 K	20 / 2 K	20 / 2 K	50 / 4 K
Задержка: размер фрейма / упрежд. буфер	20 мс / 0 мс	20 мс / 5 мс	20 мс / 5 мс	20 мс / 5 мс	20 мс / 5 мс	40 мс / 10 мс
Тип специф.: фикс. запятая	До бита	До бита	Поток бит	Поток бит	Поток бит	Поток бит

СОДЕРЖАНИЕ

Св. план 2004, поз. 15

Учебное издание

Петровский Александр Александрович,
Петровский Алексей Александрович,
Лихачёв Денис Сергеевич

РЕЧЕВЫЕ ИНТЕРФЕЙСЫ ЭВС

МЕТОДИЧЕСКОЕ ПОСОБИЕ

для студентов специальности
«Электронные вычислительные средства»
дневной формы обучения

Редактор Т.А. Лейко
Корректор Е.Н. Батутчик

Подписано в печать Формат 60x84 1/16. Бумага офсетная.
Гарнитура Times New Roman. Печать ризографическая. Усл. печ. л. .
Уч.- изд. л. Тираж 200 экз. Заказ .

Издатель и полиграфическое исполнение:
Учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники»
Лицензия ЛП № 156 от 30.12. 2002.
Лицензия ЛВ № 509 от 03.08. 2001.
220013, Минск, П. Бровки, 6.