Учреждение образования
«Высший государственный колледж связи»
Факультет электросвязи

Кафедра Программное обеспечение сетей телекоммуникаций

КУРС: ЦИФРОВАЯ ОБРАБОТКА РЕЧИ И ИЗОБРАЖЕНИЯ

для специальности:

1-45 01 03 **Сети телекоммуникаций**

Лекция № 6. Сжатие речевых сигналов на основе психоакустической модели слухового анализатора человека (2 часа)

Подготовил:
ст.преподаватель кафедры ПОСТ
Киркоров С.И.
на основе лекции Борискевича А.А.

Минск 2010

Лекция № 6. Сжатие речевых сигналов на основе психоакустической модели слухового анализатора человека (2 часа).

6.1 Психоакустические принципы сжатия аудиосигналов

Человек в состоянии принимать огромные потоки информации. Но сознательно он способен обрабатывать лишь около 100 бит/с информации. Поэтому можно говорить о присущей аудиосигналам существенной избыточности. Важной проблемой при цифровом представлении звуковых сигналов является сокращение имеющейся в них статистической и психоакустической (перцептивной) избыточности.

Сокращение статистической избыточности основывается на учете свойств самих звуковых сигналов. Она обусловлена наличием корреляционной связи между соседними отсчетами звукового сигнала при его дискретизации. Устранение статистической избыточности не приводит к значительному уменьшению скорости цифрового потока. Наиболее перспективными с этой точки зрения оказались методы сокращения психоакустической избыточности звуковых сигналов, учитывающие такие свойства слухового восприятия. Степень учета психоакустических закономерностей слухового восприятия определяет качество систем кодирования со сжатием цифровых данных. Методами устранения психофизической избыточности можно обеспечить сжатие цифровых аудиоданных в 10 – 12 раз без существенных потерь в качестве.

Описание характеристик любой слуховой системы осуществляется по частотно-пороговым кривым, которые отражают предельные значения интенсивности (или мощности звукового давления) в зависимости от частоты звуковых колебаний. Частотно-пороговые кривые, полученные для слуховой системы в целом, получили название кривых слышимости. Чувствительность слуха оценивается минимальной интенсивностью звука, при которой человек может отличить звук от постоянно существующего фона собственных шумов. Интенсивность, при которой звук обнаруживается с вероятностью 0,5 (в 50% случаев), называется порогом слышимости или абсолютным порогом для данного сигнала.

Кривая порога слышимости (рис. 6.1), характеризующая наименьшую интенсивность звука определенной частоты, который может быть услышан человеком в тишине, хорошо аппроксимируется с помощью нелинейной функции

$$T(f) = 3{,}64(f/1000)^{-0{,}8} - 6{,}5e^{-0{,}6(f/1000-3{.},3)^2} + 10^{-3}(f/1000)^4, \quad (6.1)$$

где $T(f)$ – порог слышимости, дБ; f – частота звукового сигнала, Гц. Данная кривая представляет собой геометрическое место точек, которые соответствуют тональным компонентам разных частот, имеющих одинаковую громкость.

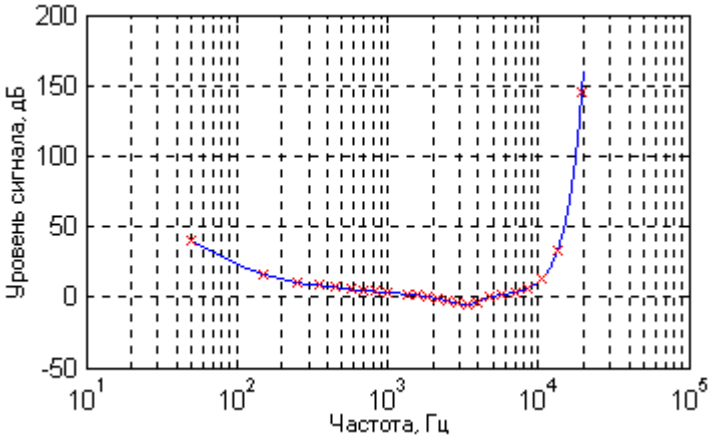


Рис. 6.1. Кривая минимального порога слышимости слухового аппарата человека в тишине

Анализ усредненного графика порога слышимости показывает, что слуховая система человека чувствительна к частотам в диапазоне 0,01–20 кГц. Наибольшая чувствительность у человека наблюдается на средних частот 2–5 кГц. Порог слышимости сигнала на частоте около 3 кГц составляет приблизительно 0 дБ. Порог слышимости значительно повышается в области низких (менее 200 Гц) и высоких (более 10 кГц) частот и может достигать значения громкости звука 40 дБ и выше. Следовательно, для слабых сигналов амплитудно-частотная характеристика слуха является нелинейной, и, как следствие, слабые низкочастотные и высокочастотные составляющие сигнала могут быть неслышны. Кривые слышимости позволяют оценить работоспособность слуховой системы по ряду признаков: диапазону воспринимаемых частот, полосе частот, в пределах которой чувствительность максимальна, абсолютным значениям звукового давления и их изменениям в зависимости от частоты.

Одно из важных свойств слухового аппарата человека является его интегрирующая способность, то есть группирование частотных составляющих звука в определенные частотные критические полосы. Слуховая система сравнивает полезный сигнал и мешающий шум по интенсивности в пределах критических полос слуха, оценивая порог слышимости. Критическая полоса слуха является полосой, за пределами которой субъективные ощущения звука сильно изменяются.

На рисунках схематически показан метод определения ширины критической полосы слуховой системы.

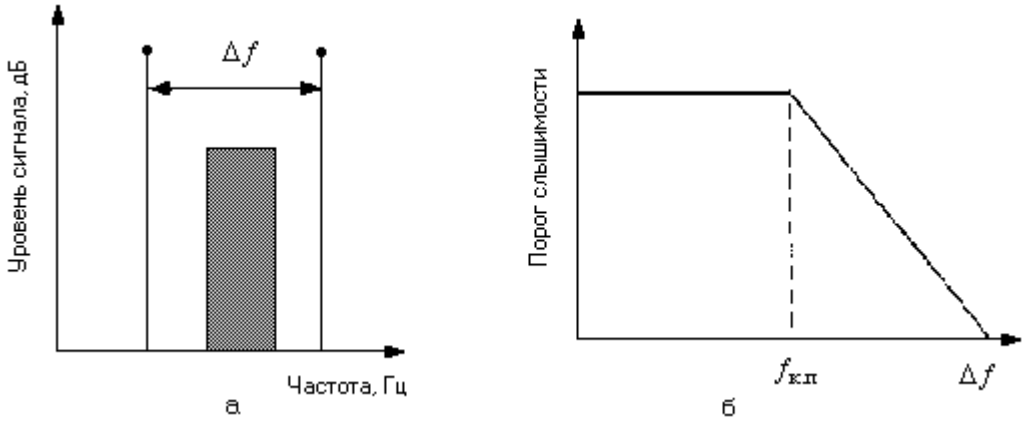


Рис. 6.2. Определение ширины критической полосы при маскировании узкополосного шумового компонента двумя тональными компонентами

Узкополосный шумовой сигнал маскируется двумя близко расположенными тональными компонентами, расстояние Δf между которыми постепенно увеличивается. На рис. 6.2,б представлена зависимость поведения порога обнаружения шумовой компоненты от изменения расстояния между двумя тональными компонентами Δf. Данный порог остается постоянным до тех пор, пока частотное расстояние между тональными компонентами остается в пределах критической полосы $f_{к.п}$. Вне полосы порог быстро уменьшается.

На рис. 6.3 приведен случай, когда в качестве маскируемого сигнала выступает тональная компонента, а в качестве маскирующих сигналов – две узкополосные шумовые компоненты (рис. 6.3,а). Поведение порога обнаружения для тональной компоненты (рис. 6.3,б) аналогично рис. 6.2,б.

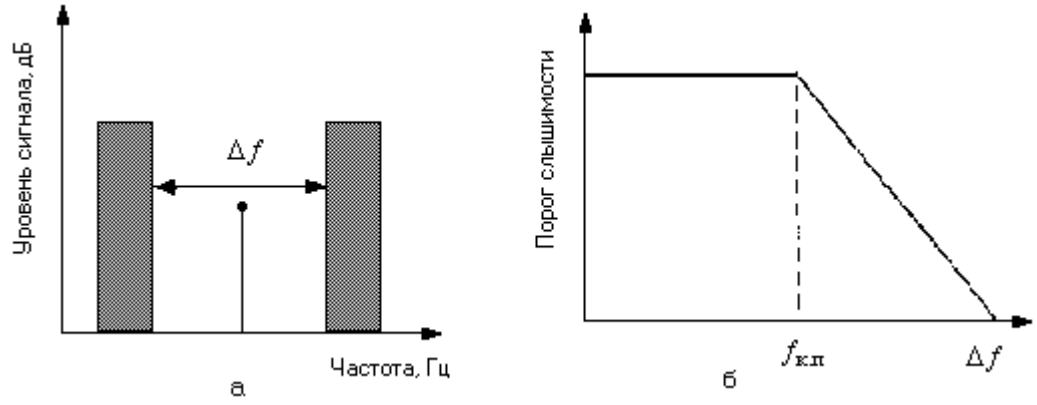


Рис. 6.3. Определение ширины критической полосы при маскировании тонального компонента двумя узкополосными шумовыми компонентами

В табл. 6.1 представлено разбиение слышимого диапазона частот на критические полосы с соответствующими центральной частотой и шириной полос идеализированного банка фильтров с прямоугольными амплитудно-частотными характеристиками, который позволяет качественно воспроизвести звуковые сигналы. Из таблицы видно, что критическая полоса

не соответствует диапазону с фиксированными нижней и верхней границами. Ширина критической полосы остается приблизительно постоянной (около 100 Гц) вплоть до 500 Гц и увеличивается приблизительно на 20% от центральной частоты при 500 Гц и выше:

$$BW(f) = \begin{cases} 100\ Гц, & f < 500\ Гц, \\ 0,2\ fГц\ , & f \geq 500Гц\ . \end{cases} \qquad (6.2)$$

Из выражения видно, что в частотной области до 500 Гц органы слуха воспринимают интенсивность звука, разделяя ее на участки постоянной абсолютной ширины, а свыше 500 Гц – на участки постоянной относительной ширины.

Для среднего слушателя ширина критической полосы слуховых фильтров может быть определена из выражения

Таблица 6.1. Разбиение слышимого частотного диапазона на критические полосы в психоакустической модели слуховой системы

Номер критической полосы	Центральная частота, Гц	Нижняя и верхняя частоты, ширина критической полосы, Гц	
1	50	20–100	100
2	150	100–200	100
3	250	200–300	100
4	350	300–400	100
5	450	400–510	110
6	570	510–630	120
7	700	630–770	140
8	840	770–920	150
9	1000	920–1080	160
10	1170	1080–1270	190
11	1370	1270–1480	210
12	1600	1480–1720	240
13	1850	1720–2000	280
14	2150	2000–2320	320
15	2500	2320–2700	380
16	2900	2700–3150	450
17	3400	3150–3700	550
18	4000	3700–4400	700
19	4800	4400–5300	900
20	5800	5300–6400	1100
21	7000	6400–7700	1300
22	8500	7700–9500	1800
23	10500	9500–12000	2500
24	13500	12000–15500	3500
25	19500	15500–22500	7000

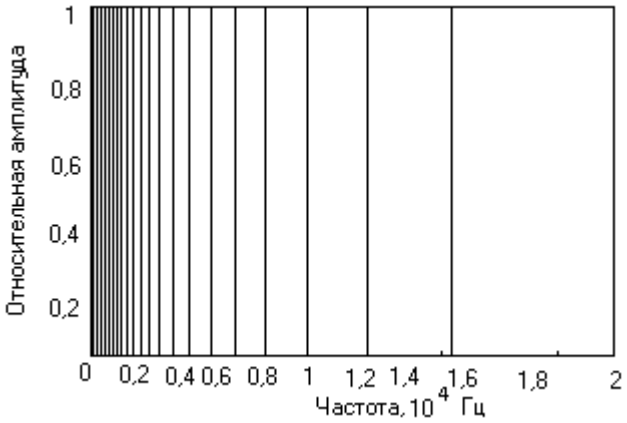


Рис. 6.4. Идеализированный банк фильтров как упрощенная модель слуховой системы

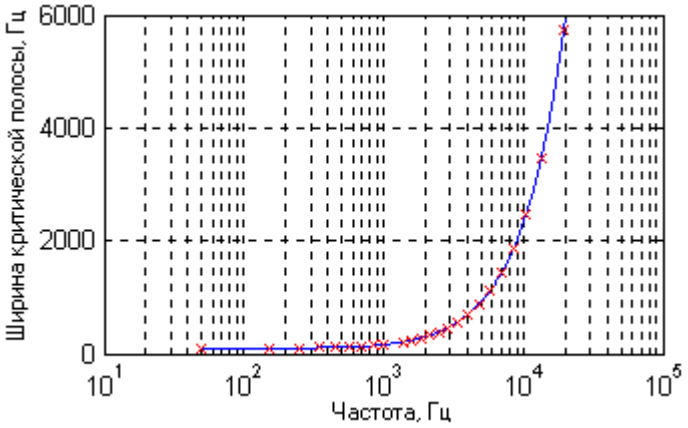


Рис. 6.5 . Ширина критических полос как функция центральной частоты

$$BW(f) = 25 + 75\left[1 + 1{,}4(f/1000)^2\right]^{0{,}69} . \qquad (6.3)$$

Из рис. 6.5. видно, что по мере увеличения центральной частоты критическая полоса расширяется. Поскольку имеет место однозначная зависимость между частотой синусоидального колебания и высотой тона, то по аналогии с частотной группой слуха было введено понятие тональных групп слуха. Ширина тональной группы не зависит от высоты тона и составляет 100 мел (один мел равен ощущаемой высоте звука частотой 1000 Гц при уровне интенсивности 40 дБ), поэтому оказалось удобным ввести единицу высоты тона в одну тональную группу слуха. В честь немецкого физика Г. Баркгаузена она носит название барк (1 барк = 100 мел), равная ширине критической полосы или расстоянию между центральными частотами соседних критических полос. Переход от одной полосы к другой соответствует изменению высоты в 1 барк.

Функция для перехода от частоты в Гц к частоте в барках (рис. 6.7) определяется соотношением

$$z(f) = 13\,arctg(0{,}00076\,f) + 3{,}5\,arctg\left[\left(\frac{f}{7500}\right)\right]^2 , \qquad (6.4)$$

где $z(f)$ представляет собой частоту f в барках.

В основе способов выделения в звуковом сигнале слышимой части информации и удаления из сигнала всех остальных неслышимых деталей лежит эффект маскирования. Он связан с процессом взаимодействия сигналов, приводящим к изменению слуховой чувствительности к маскируемому сигналу в присутствии маскирующего.

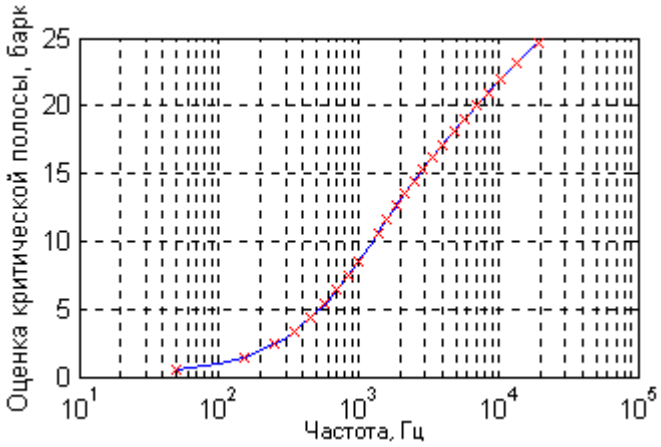


Рис. 6.6. Расстояние между центральными частотами критических полос

Эффекты слухового маскирования зависят от спектральных и временных характеристик маскируемого сигнала и сигнала маскирования и могут быть разделены на две основные группы: частотное (одновременное) и временное (неодновременное) маскирования. Маскирование по частоте заключается в следующем: если два сигнала одновременно находятся в ограниченной частотной области, то более слабый сигнал становится неслышимым на фоне более сильного. Маскирование по времени определяет следующий эффект: более слабый сигнал становится неслышимым за 5 – 20 мс до включения сигнала маскирования и становится слышимым через 50 – 200 мс после его включения.

Маскирование непосредственно связано с нелинейностью слуха. Она проявляется в том, что при воздействии на барабанную перепонку достаточно громкого синусоидального звука с частотой f в слуховом аппарате формируются гармоники этого звука с частотами $2f$, $3f$ и т.д. Поскольку в первичном воздействующем тоне этих гармоник нет, они получили название субъективных гармоник.

Анализ кривых маскировки (кривых порога слышимости при наличии маскировки) показывает, что маскировка сигналов, имеющих частоту, лежащую ниже частоты мешающего тона, проявляется значительно слабее чем вышележащих частот. Это дает основание предполагать, что маскирующее действие обусловлено возникновением субъективных гармоник мешающего тона.

Если в качестве мешающего звука использовать узкополосный белый шум, пороги слышимости определяются более четко, а сами кривые получаются более симметричными. Это связано с отсутствием

гармонических биений. Большой практический интерес представляют кривые порога слышимости, полученные при маскировке широкополосными шумами.

На рис. 6.7 приведены кривые порогов слышимости синусоидальных звуков, полученные при маскировке белым шумом с различной спектральной плотностью. Из рис. 6.7 видно, что на низких частотах кривые практически не зависят от частоты (примерно до 500 Гц). При дальнейшем увеличении частоты уровень порога слышимости повышается на 3 дБ при каждом удвоении частоты, что примерно соответствует зависимости расширения ширины критических полос с увеличением центральной частоты.

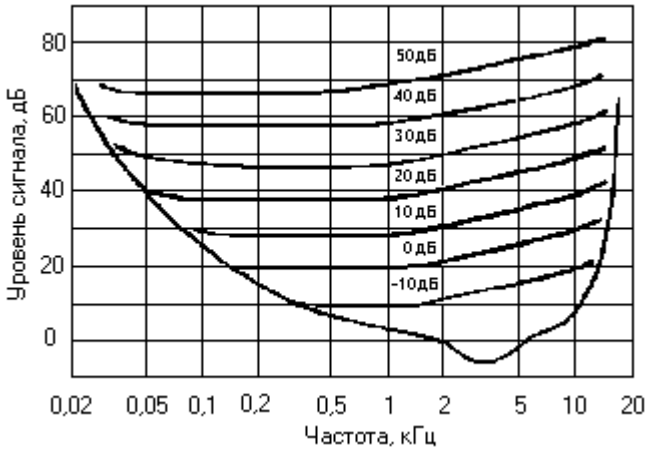


Рис. 6.7. Кривые порога слышимости при маскировке белым шумом

Из этого следует, что белый шум неодинаково эффективен для маскировки разных частот. Свойство широкополосных сигналов оказывать максимальное влияние на маскировку сигнала в пределах критических полос положено в основу современных психоакустических алгоритмов сжатия сигналов.

Эффект маскирования в частотной области связан с тем, что в присутствии больших звуковых амплитуд человеческое ухо нечувствительно к малым амплитудам близких частот (рис. 6.8).

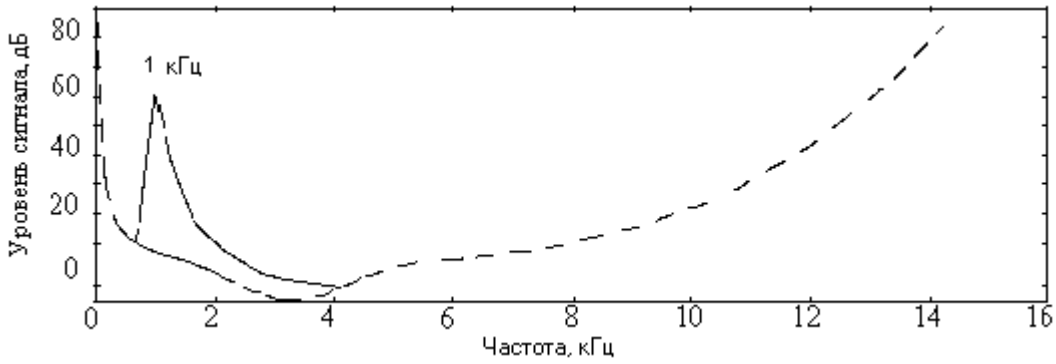


Рис.6.8. Изменение порога чувствительности слуховой системы при маскировании тональным сигналом определенной частоты

Таким образом, для маскирования важнее не абсолютная громкость звуков, а отношение мощности громкого сигнала к мощности тихого. Кроме того, чем ближе частота к частоте маскирующего сигнала, тем сильнее эффект маскирования (рис. 6.8). Степень маскировки – это разность в децибелах между уровнем порога слышимости маскируемого тона в присутствии маскера и его уровнем порога слышимости в тишине. На рис. 6.8 показаны зависимость абсолютного порога чувствительности уха и характер его изменения в присутствии тона с частотой 1 кГц и амплитудой в 60 дБ. Из рис. 6.8 видно, что порог слышимости имеет адаптивный характер: он повышается при появлении каких-либо звуков и понижается в тишине. Для того, чтобы услышать более слабые звуки на фоне более сильных, их частоты должны значительно различаться. Кривая порога чувствительности во многом подобна изменяющейся во времени функции кратковременной спектральной плотности мощности и имеет локальные максимумы в соответствии с тональными компонентами высокого уровня.

Из рис. 6.9 видно, что с увеличением средней частоты маскирующего сигнала диапазон частот, где проявляется маскировка, становится шире. Кривые маскировки являются несимметричными, они имеют крутой спад в сторону низких частот и пологий спад в сторону высоких частот.

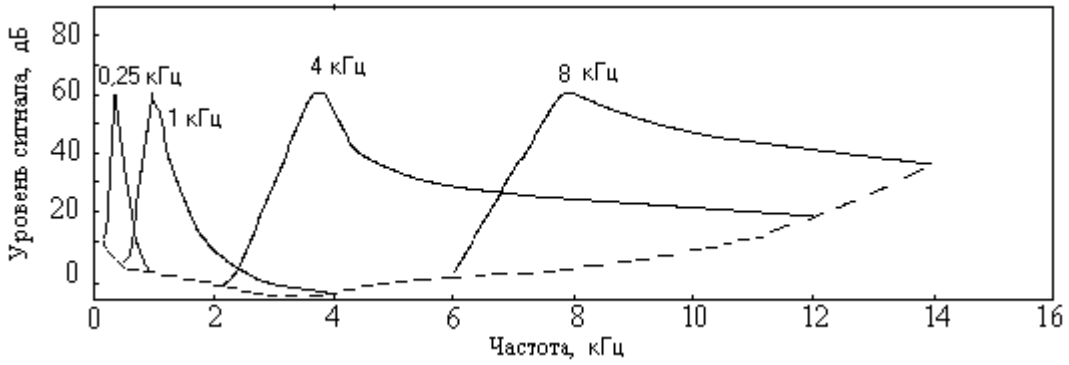


Рис. 6.9. Изменение порога чувствительности слуховой системы при маскировании сигналами со средними частотами полос 0,25, 1, 4 и 8 кГц и уровнем интенсивности 60 дБ

Если при построении кривых маскировки по оси абсцисс откладывать не частоты f, а значения высоты тона z в барках, то кривые маскировки при разных значениях z маскирующего сигнала будут одинаковыми при одном и том же значении уровня маскирующего сигнала (рис.6.10). Их форма при таком представлении будет зависеть только от маскирующего сигнала, и не будет зависеть от величины z.

Маскирование имеет место не только внутри критической полосы, оно распространяется и на соседние полосы. Маскирующий сигнал, центрированный в пределах критической полосы, оказывает определенное воздействие на порог обнаружения в соседних критических полосах. Функция протяженности маскирования, которая учитывает маскирование как внутри, так и между полосами, задается аналитическим выражением

$$SF(x) = 15{,}81 + 7{,}5(x + 0{,}474) - 17{,}5\sqrt{1 + (x + 0{,}474)^2}\,, \qquad (6.5)$$

где x и $SF(x)$ представлены соответственно в барках и дБ.

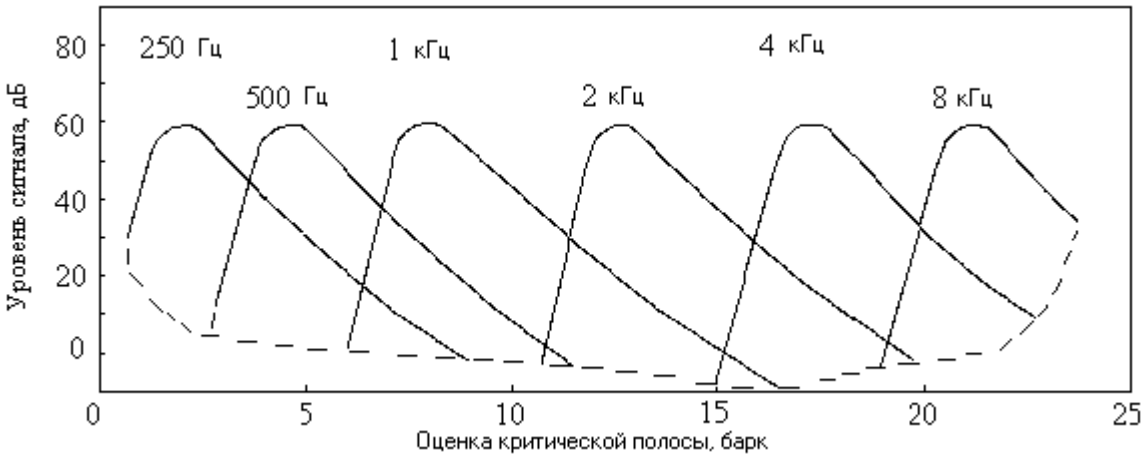


Рис. 6.10. Частотное маскирование с частотной шкалой в барках

Вследствие разделения спектра звукового сигнала на частотные полосы слух реагирует не на общую мощность сигнала, а на мощность, сосредоточенную в отдельных частотных группах. При этом более интенсивные частотные группы при определенных условиях могут маскировать менее интенсивные группы. Применительно к сжатию это означает, что в группах меньшей интенсивности допустим больший уровень шума квантования, т.е. допустимо уменьшение числа разрядов при кодировании соответствующих спектральных составляющих, что в свою очередь снижает скорость передаваемого цифрового потока.

Маскирование во временной области (рис. 6.11) характеризует динамические свойства слуха, показывая изменение во времени относительного

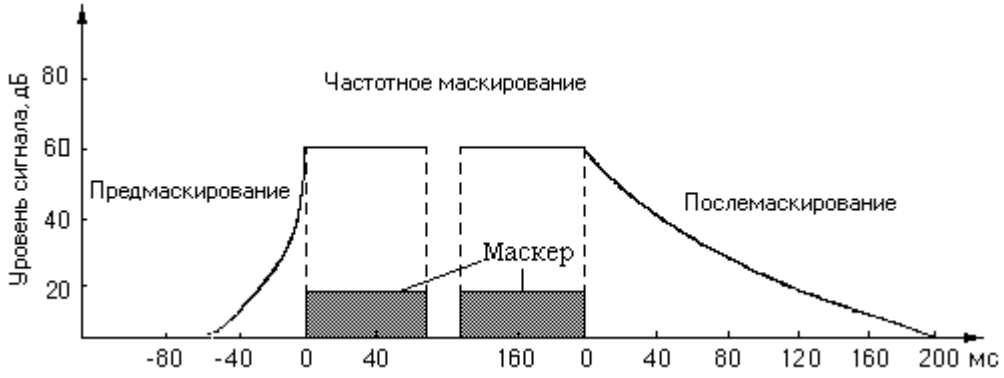


Рис. 6.11. Зависимость изменения порога слышимости маскируемого сигнала от величины временного интервала между сигналом и маскером

порога слышимости (порог слышимости одного сигнала в присутствии другого), когда маскирующий и маскируемый сигналы звучат не одновременно.

Различают предмаскировку (маскируемый сигнал перед маскером) и послемаскировку (маскируемый сигнал после маскера). Оба вида маскировки наступают, если маскируемый сигнал и маскер одновременно находятся во временном окне шириной 100–200 мс. Считается, что за это время завершается полная обработка сигнала с выделением всех его признаков. В этом случае слуховую систему можно рассматривать как процессор для обработки звуковых сигналов во временном окне шириной примерно 200 мс.

Из рис. 6.11 видно, что длительность типичных промежутков времени, в пределах которых действуют предмаскирование и послемаскирование, составляет соответственно 5 – 50 мс и 50 – 200 мс. Предмаскировка проявляется на значительно более коротком временном интервале, причем ее длительность в очень сильной степени зависит от опыта слушателя. Степень маскирования звукового сигнала зависит от интервала между маскируемым сигналом и маскером, уровня интенсивности и длительности воздействия маскера. Сближение во времени подачи сигнала и маскера увеличивает маскировку. Временное маскирование зависит от частотного взаимоотношения сигнала и маскера точно так же, как и при частотном маскировании. Оно проявится в большей степени, если сигнал и маскер близки по частоте.

6.2 Психоакустическая модель слуховой системы в частотной области

Алгоритм работы психоакустической модели слуховой системы предназначен для вычисления глобальных маскирующих порогов и отношения сигнал/маска и состоит из следующих пяти этапов.

1.Спектральный анализ и нормализация уровня звукового давления сигнала. Выборки (отсчеты) входного аудиосигнала *s(n)* нормализуются по формуле

$$x(n)=\frac{s(n)}{N(2^{b-1})}, \tag{6.6}$$

где N – количество отсчетов или длина окна анализа сигнала; b – количество разрядов (битов) квантования на выборку; n – индекс выборки входного сигнала.

Нормализованный сигнал $x(n)$, используя оконную функцию Ханна с 1/16 долей перекрытия, разбивается на блоки, содержащие 512 выборок входного сигнала. Оценка спектральной плотности мощности $P(k)$ в дБ с использованием 512-точечного быстрого преобразования Фурье производится по формуле

$$P(k)=PN+10\log_{10}\left|\sum_{n=0}^{N-1}w(n)x(n)e^{-j\frac{2\pi kn}{N}}\right|^2 \quad \text{при} \quad 0\le k\le\frac{N}{2}, \tag{6.7}$$

где PN – уровень нормализации мощности сигнала, равный 90 дБ; k – индекс спектральной составляющей интервала разрешения (бина); $w(n)=\frac{1}{2}\left[1-\cos\left(\frac{2\pi n}{N}\right)\right]$ – оконная весовая функция Ханна.

Поскольку уровни воспроизведения сигнала неизвестны в момент проведения психоакустического анализа сигнала, процедура нормализации и параметр PN используются для оценки уровня звукового давления независимо от входного сигнала.

2.Идентификация тональных и шумовых маскеров. Локальные максимумы в спектральной плотности мощности $P(k)$, которые превышают соседние компоненты в пределах одного барка по крайней мере на 7 дБ, идентифицируются как тональные. Множество тональных спектральных максимумов S_T определяется следующим выражением:

$$S_T=\left\{P(k)\left|\begin{array}{l}P(k)>P(k\pm1),\\P(k)>P(k\pm\Delta_k)+7dB\end{array}\right.\right\}, \tag{6.8}$$

где $P(k)$ – $к$-й локальный спектральный максимум;

$$\Delta_k\in\begin{cases}2 & при \quad 2<k<63 \quad (0.17-5.5\ кГц),\\ [2,3] & при \quad 63\le k<127 \quad (5.5-11\ кГц),\\ [2,6] & при \quad 127\le k\le256 \quad (11-20\ кГц).\end{cases}$$

Значения тональных маскеров $P_{TM}(k)$ в дБ вычисляются из спектральных максимумов множества (4.81) по формуле

$$P_{TM}(k) = 10\log_{10}\sum_{j=-1}^{1} 10^{0,1P(k+j)}, \qquad (6.9)$$

где $P(k-1)$ и $P(k+1)$- значения спектральной плотности мощности частотных составляющих, соседних с $P(k)$.

После идентификации тональных маскеров значение спектральной плотности мощности $P_{NM}(\bar{k})$ в дБ единственного шумового маскера для каждой критической полосы вычисляется из оставшихся спектральных составляющих, находящихся за пределами диапазона тонального маскера $\pm\Delta_k$, по формуле

$$P_{NM}(\bar{k}) = 10\log_{10}\sum_{j} 10^{0,1P(j)} \qquad (6.10)$$

для всех $P(j) \notin \{P_{TM}(k, k\pm 1, k\pm\Delta_k)\}$, где $\bar{k} = \left(\prod_{j=1}^{u} j\right)^{1/(l-u+1)}$ – средняя геометрическая спектральная составляющая критической полосы; l и u – нижняя и верхняя границы спектральных составляющих критической полосы соответственно.

3.Прореживание и преобразование маскеров. На этом шаге количество маскеров сокращается на основе двух критериев: абсолютного порога слышимости и частотного прореживания. Тональные и шумовые маскеры, расположенные ниже абсолютного порога слышимости, отбрасываются, а остаются лишь маскеры, удовлетворяющие условию

$$P_{TM,NM}(k) \geq T_q(k), \qquad (6.11)$$

где $T_q(k)$ – уровень звукового давления абсолютного порога слышимости в тишине для k-й спектральной составляющей.

Процедура скользящего окна шириной 0,5 барка используется для замены любой пары маскеров, появляющихся в пределах данного окна, одним маскером с наибольшей мощностью. После выполнения данной процедуры частотные бины (интервалы разрешения) маскеров преобразуются согласно схеме субдискретизации

$$P_{TM,NM}(i) = P_{TM,NM}(k), \qquad (6.12)$$

$$P_{TM,NM}(k) = 0, \qquad (6.13)$$

где $i = \begin{cases} k & 1 \leq k \leq 48, \\ \text{при} & 49 \leq k \leq 96, \\ k + & 97 \leq k \leq 232. \end{cases}$ (6.14)

Из соотношения (6.14) следует, что эффект прореживания бинов маскеров в критических полосах 18–22 слуховой системы составляет 2:1, а для критических полос 22–25 он равен 4:1 без потерь маскирующих компонент. Данная процедура уменьшает общее число частотных бинов тональных и шумовых маскеров от 256 до 106. Результатом выполнения данного этапа является прореженное множество тональных и шумовых маскеров.

Данный этап удаляет спектральные компоненты, которые присутствуют обычно ниже порога слышимости в тишине, и более слабые тональные компоненты, находящиеся в пределах половины ширины критической полосы (0,5 барка) сильного тонального компонента. Данная модель использует прореженные тональные и нетональные (шумовые) компоненты для определения глобального порога маскирования в частотной области.

4.Вычисление индивидуальных порогов маскирования. Каждый индивидуальный порог характеризует маскирующий вклад в *i*-й частотный бин тонального или шумового маскера, расположенного в *j*-м бине, преобразованном на этапе 3. Индивидуальные пороги тонального маскера $T_{TM}(i,j)$ вычисляются по формуле

$$T_{TM}(i,j) = P_{TM}(j) - 0{,}275z(j) + SF(i,j) - 6{,}025, \qquad (6.15)$$

где $P_{TM}(j)$ – уровень звукового давления тонального маскера в дБ в *j*-м частотном бине; $z(j)$ – частота *j*-го бина в барках; $SF(i,j)$ – функция распространения маскирования от *j*-го бина маскера к *i*-му бину маскируемого сигнала.

Функция распространения маскирования $SF(i,j)$ в дБ от *j*-го бина маскера к *i*-му бину маскируемого сигнала моделируется в виде кусочно-линейной функции уровня маскера $P(j)$ по формуле

$$SF(i,j) = \begin{cases} 17\Delta_z - 0{,}4P_{TM}(j) + 11 & -3 \le \Delta_z < -1, \\ (0{,}4P_{TM}(j) + 6)\Delta_z & -1 \le \Delta_z < 0, \\ -17\Delta_z & при \quad 0 \le \Delta_z < 1, \\ (0{,}15P(j)_{TM} - 17)\Delta_z - 0{,}15P(j)_{TM} & 1 \le \Delta_z < 8, \end{cases} \qquad (6.16)$$

где $\Delta_z = z(i) - z(j)$ – расстояние в барках для разделения маскера и маскируемого сигнала. В данной модели для увеличения вычислительной эффективности протяженность маскирования ограничивается окрестностью шириной до 10 барк.

Индивидуальные пороги шумового маскера $T_{NM}(i,j)$вычисляются по формуле

$$T_{NM}(i,j) = P_{NM}(j) - 0{,}175z(j) + SF(i,j) - 2{,}025, \quad (6.17)$$

где $P_{NM}(j)$ – уровень звукового давления шумового маскера в дБ в j-м частотном бине; $z(j)$ – частота j-го бина в барках. Функция $SF(i,j)$ вычисляется путем замены $P_{TM}(j)$ на $P_{NM}(j)$ в соотношении (6.15).

5. Вычисление глобальных порогов маскирования. Индивидуальные пороги маскирования объединяются, чтобы оценить глобальный порог маскирования для каждого частотного бина в подмножестве. Данная модель основана на свойстве аддитивности маскирующих эффектов. Глобальный порог маскирования $T_g(i)$ вычисляется следующим образом:

$$T_g(i) = 10\log_{10}\left(10^{0{,}1T_q(i)} + \sum_{l=1}^{L} 10^{0{,}1T_{TM}(i,l)} + \sum_{m=1}^{M} 10^{0{,}1T_{NM}(i,m)}\right), \quad (6.18)$$

где $T_q(i)$ – абсолютный порог слышимости для i-го частотного бина; $T_{TM}(i,l)$ и $T_{NM}(i,l)$ – индивидуальные пороги маскирования, вычисленные на этапе 4; L и M – количество соответственно тональных и шумовых маскеров, идентифицируемых на этапе 3.

Таким образом, глобальный маскирующий порог для каждого частотного бина представляет собой сигнало-зависимую, аддитивную по мощности модификацию абсолютного порога слышимости в тишине. В этом случае тональная кривая маскирования, нетональная (шумовая) кривая маскирования и кривая абсолютного порога складываются для получения общей кривой маскирования анализируемого входного сигнала. Таким образом, глобальный маскирующий порог определяется сложением всех индивидуальных порогов и порога в тишине. Добавление последнего порога гарантирует, что глобальный маскирующий порог будет не ниже порога слышимости. Маскирующий эффект в каждой критической полосе слуховой системы характеризуется величиной отношения сигнал/маска. В данной модели отношение сигнал/маска определяется как отношение максимума мощности сигнала в пределах критической полосы и глобального маскирующего порога, при котором шум квантования еще маскируется полезным сигналом и не замечается слуховым анализатором человека.

Технологии сжатия, основанные на использовании психоакустических моделей, имеют один общий недостаток: все они работают качественно до скорости битового потока (битрейт) 128 Кбит/с при частоте дискретизации 44.1 кГц с разрешением 16 разрядов на отсчет. Нормальный для слухового восприятия битрейт составляет 320 Кбит/с. Битрейт характеризует степень сжатия аудиоданных. Чем меньше эта величина, тем больше степень сжатия и тем ниже качество звучания. В связи с этим был разработан кодек MP3 Pro, который является развитием MP3 и использует новую технологию SBR (Spectral Band Replication) для использования битрейта 64 Кбит/с. Данная

технология дополняет использование психоакустической модели и предназначена для передачи верхнего частотного диапазона. Сущность технологии SBR заключается в том, что кодируется чуть более узкий диапазон частот с обрезанными верхними частотами, которые восстанавливаются декодером на основе информации о более низких частотных составляющих. Таким образом, слышимый сигнал представляет собой уже не столько оригинал, сколько синтезированную копию оригинала.

6.3 Сжатие аудиосигнала на основе стандарта MP3

Международная организация по стандартизации (International Organization for Standardization - ISO) и экспертная группа по вопросам движущего изображения (Motion Picture Experts Group - MPEG) разработали стандарт аудиосжатия для сигнала, синхронизированного с сжатым видеосигналом, известный как MPEG. В стандарте кодирования аудиосигналов MPEG-1 использованы три уровня (Layer 1, 2 и 3) увеличивающей сложности и улучшающей субъективной производительности, частоты дискретизации 32, 44,1 и 48 кГц, а биты на выход подаются со скоростью от 32 до 192 Кбит/с (монофонический канал) или со скоростью от 64 до 384 Кбит/с (стереофонический канал). При данных высоких значениях скорости потока двоичных данных каждый из уровней стандарта MPEG-1 обеспечивает приблизительно одинаковое качество звучания, близкое к качеству компакт-диска. Система сжатия звука MPEG-1 Layer 3 (MP3) позволяет регулировать коэффициент сжатия и искать оптимальное соотношение качество звучания/коэффициент сжатия.

Перцептивный кодер, используемый в стандарте MPEG, является примером эффективного сжатия аудиоинформации с учетом особенностей слуховой системы человека. На рисунке представлена структурная схема перцептивного кодера на основе MP3. Из рис. 6.12 видно, что цифровой 16-битовый аудиосигнал разделяется на M = 32 частотные полосы с помощью группы равномерно расположенных полосовых фильтров. Все фильтры в группе фильтров получаются путем модулирования фильтра прототипа, предназначенного для увеличения частотного разрешения аудиосигнала и лучшей аппроксимации критических полос слуховой системы. Импульсный отклик фильтра *i*-й полосы на единичный импульс $h_i(k)$ можно представить в виде произведения коэффициентов $h(k)$ фильтра прототипа, являющихся квантованными значениями импульсной характеристики этого фильтра, и модулирующей функции, которая сдвигает низкочастотный отклик в *i*-ю частотную полосу,

$$h_i(k) = h(k)\cos\left(\frac{2i-1}{2M} + \phi(i)\right), \qquad (6.19)$$

где M – число полос, на которые разделяется звуковой сигнал $x(t)$; i=1,2,.., 32; k – номер коэффициента импульсной характеристики $h(k)$ фильтра прототипа (k=1,2,…512), $\phi(i)$ - фазовый сдвиг для i-й полосы.

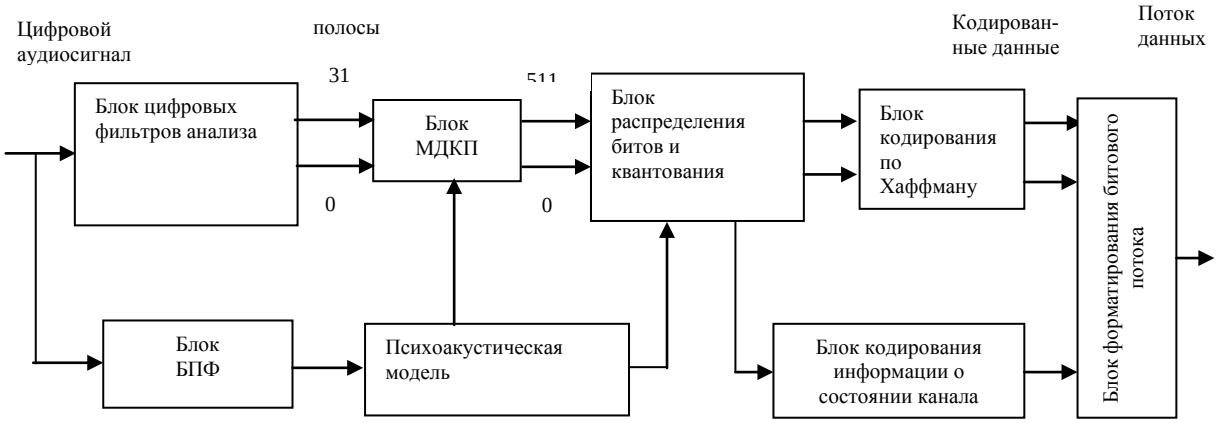


Рис. 6.12. Структурная схема перцептивного кодера MP3

Сигнал на выходе фильтра i-й полосы в дискретный момент времени n представляется в виде дискретной свертки

$$y_i(n) = x(n) * h_i(n) = \sum_{k=0}^{N-1} x(t-k)h_i(k)\ , \qquad (6.20)$$

где * – символ свертки; $h_i(k)$ – массив коэффициентов i-го полосового фильтра порядка N=512; $x(n-k)$– входной массив цифровых данных фильтра. Число N определяет эффективность фильтра, реализующего N операций умножения с накоплением для каждого выходного отсчета $y_i(n)$. Подход к кодированию, основанный на использовании набора цифровых фильтров, дает высокое временное разрешение и таким образом, допускает более точное моделирование эффекта маскирования. Полосы нормализуются с помощью масштабирующего фактора так, что максимальное значение амплитуды отсчета в каждой полосе равно единице. Повышение точности распознавания маскирующих сигналов достигается посредством дальнейшей обработки 32 полосовых сигналов с помощью перекрывающегося модифицированного дискретно косинусного преобразования (МДКП), предназначенного для анализа соседних областей частотных полос. Это преобразование задается выражением

$$Y(n) = \sum_{k=0}^{N-1} h(k)y(k)\cos\left(\frac{\pi}{2N}(2n+1)(2k+1+\frac{N}{2})\right), \qquad (6.21)$$

где n и k – индексы спектральной компоненты $Y(k)$ и дискретного отсчета $y(k)$ аудиосигнала соответственно; $n=0,1, \ldots, \frac{N}{2}-1$, $h(k)-$ импульсный отклик окна анализа сигнала $x(k)$, обладающий свойствами фильтра нижних частот. МДКП основано на идеи устранения наложения во временной области путем использования пересекающих временных окон анализа полосовых сигналов. Наложение во временной области, вводимое с помощью 50% перекрытия окон, устраняется при обратном преобразовании МДКП. Для более точного восстановления аудиосигнала весовые коэффициенты $h(k)$ функции анализирующего окна должны удовлетворять соотношениям

$$h^2\left(\frac{N}{2}-1-k\right)+h^2(k)=2, \qquad 0 \le k \le N/2, \tag{6.22}$$

$$h^2\left(\frac{N}{2}+k\right)+h^2(N-1-k)=2, \qquad 0 \le k \le N/2. \tag{6.23}$$

С практической точки зрения ни сигналы, ни реализуемые узкополосные фильтры не имеют строго ограниченной полосы. Следовательно, всегда имеет место некоторое перекрытие соседних частотных полос, что приведет к неизбежным искажениям при восстановлении непрерывного сигнала. Для устранения нежелаемого временного эффекта наложения, называемого *aliasing* (явление, являющееся результатом недостаточной частоты выборки и сопровождаемое появлением ложных частот), окна анализа задаются выражением

$$h(k)=\pm\sqrt{2}\,\sin\left[\left(k+\frac{1}{2}\right)\frac{\pi}{N}\right]. \tag{6.24}$$

В перцептивном кодере часто используется синусоидальное окно $h(k)=\sin\left(\pi\left(\frac{k}{N}\right)\right)$, также удовлетворяющее требованиям совершенного восстановления непрерывного аудиосигнала.

Кодер MP3 реализует метод адаптивный обработки значений МДКП, позволяющий устранить некоторые артефакты, вызванные перекрывающими полосами группы фильтров анализа и более эффективно использовать психоакустические свойства слуховой системы. Кодер использует два блока МДКП различной длины: длинный блок с 18 выборками и короткий блок с 6 выборками. Поскольку имеет место 50% перекрытие между последовательными окнами МДКП, то ширины окон преобразования достигают длительностей 36 и 12 выборок для длинного и короткого блоков соответственно

На рис. 6.13 приведена структурная схема блока МДКП перцептивного кодера MP3. Для изменения длины МДКП используется набор оконных функций: синусное окно типа short для длинного блока, синусное окно short для короткого блока и два окна перехода start и stop, необходимые для уменьшения искажений, возникающих при переходе от длинных к коротким окнам и наоборот. Решение о выборе длины МДКП принимает психоакустическая модель кодера. Из рис. 6.13 видно, что выходные полосовые сигналы обрабатываются с помощью МДКП в двух различных по длине блоках: коротком (16 - точечное МДКП) и длинном (36- точечное МДКП). Короткий блок улучшает временное разрешение кодера, что позволяет эффективно использовать предмаскирование для скрытия различных артефактов, возникающих при кодировании. Длинный блок позволяет
реализовать более высокое частотное разрешение для квазистационарных сигналов, что сокращает дополнительную информацию о состоянии канала и уменьшает скорость аудио потока.

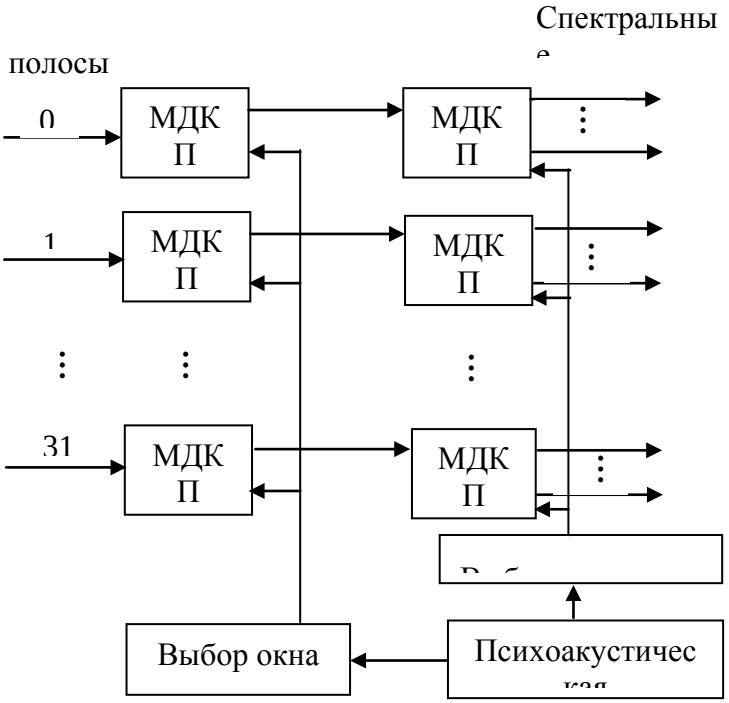


Рис. 6.13. Структурная схема блока МДКП перцептивного кодера MP3

Для заданного кадра аудио выборок МДКП может иметь все блоки одинаковой длины (длинные или короткие) или иметь режим смешанных блоков. Применение 18-точечного МДКП к выходным отсчетам одной из 32 полос анализирующего блока фильтров позволяет улучшить частотное разрешение, приблизив его к критическим полосам. Максимальное число компонент в кодере равно 32·18=576, а частотное разрешение составляет 24000/576 или 41,67 Гц.

Для спектральной декомпозиции аудиосигнала и достижения достаточно высокого спектрального разрешения при вычислении

маскирующих порогов параллельно с разбиением аудиосигнала на полосы вычисляется 1024 - точечное БПФ. Психоакустическая модель использует результаты БПФ для разделения полосовых данных на длинные и короткие блоки в зависимости от того, какое требуется частотное или временное разрешение, и обнаружения тональных и нетональных компонентов спектра. Значение тональных компонентов вычисляется путем нахождения локальных максимумов спектра. Маскирующие эффекты, обусловленные тональными и нетональными компонентами и порогом слышимости, используются при вычислении минимального маскирующего порога для каждой полосы. Далее следует итеративная процедура распределения разрядов, которая использует вычисленные маскирующие пороги для выбора оптимального квантователя для каждой частотной полосы. Квантователи выбираются таким образом, чтобы одновременно удовлетворить требованиям маскирования и заданного битрейта. Сущность разрядного (битового) распределения состоит в том, что часто определенные фрагменты аудиосигнала не могут быть закодированны в рамках данного разрядного диапазона без потерь качества. В таком случае MP3 использует небольшой запас разрядов, имеющихся в буфере, для кодирования менее сложных фрагментов в меньший разрядовый диапазон. Данный режим работы предназначен для кодирования блоков отсчетов входного сигнала, содержащих легко сжимаемые сигналы, минимально необходимым числом битов, а оставшие разряды добавляются в блоки со сложным, трудно компрессируемым сигналом.

Классический алгоритм Хаффмана в MP3 используется на последнем этапе сжатия. Это так называемое энтропийное кодирование, учитывающее статистические особенности звукового сигнала. Кодирование Хаффмана хорошо дополняет перцептивные методы сжатия. Перцептивное кодирование особенно эффективно при кодировании сложного полифонического звучания в силу того, что многие звуки маскируются и их влияние на общую картину снижается. При кодировании же чистых звуков оказывается эффективным алгоритм Хаффмана из-за того, что оцифрованный звук содержит много избыточных байтов, которые при кодировании заменяются более короткими кодами. Квантованные данные (значения коэффициентов МДКП) вместе с информацией о распределении битов, квантованных масштабирующих коэффициентах и выбранном квантователе передаются в блок временного уплотнения (мультиплексор) для передачи потока данных по каналу связи.

На рис 6.14 приведена упрощенная структурная схема перцептивного декодера, восстанавливающего аудиоданные с блочной структурой из потока данных на выходе кодера.

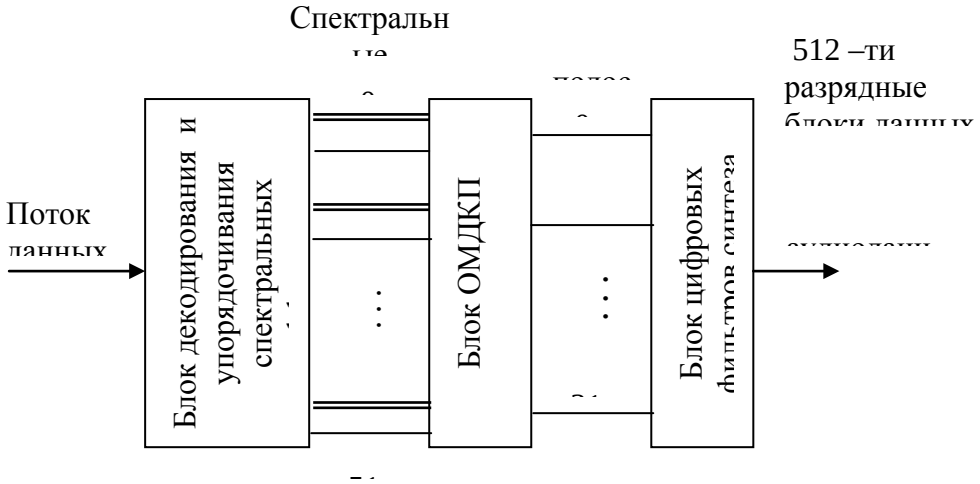


Рис. 6.14. Структурная схема перцептивного декодера MP3

В соответствии с преобразованием (6.21) обратное модифицированное дискретно-косинусное преобразование (ОМДКП) задается выражением

$$y(m) = h(m)\frac{4}{\pi}\sum_{n=0}^{\frac{N}{2}-1} Y(n)\cos\left(\frac{\pi}{2N}(2n+1)\left(2m+1+\frac{N}{2}\right)\right), \qquad (6.25)$$

где *m=0,1,…,N-1*. Обозначение y в преобразовании (6.25) подчеркивает тот факт, что последовательность данных, восстановленная с помощью МДКП, не соответствует исходной последовательности данных. Блок цифровых фильтров синтеза осуществляет обратное преобразование полос аудиосигнала, первоначально формируемых блоком цифровых фильтров анализа, в аудиосигнал с блочной структурой.

6.4. Прямое и обратное МДКП

Данное преобразование задается выражением

$$X(k) = \sum_{n=0}^{2M-1} x(n)h(n), \qquad (6.26)$$

где k и n – индексы спектральной компоненты $X(k)$ и дискретного отсчета $x(k)$ речевого сигнала соответственно; $0 \le k \le M-1$; $h_k(n) = w(n)\sqrt{2/M}\cos\left[\frac{(2n+M+1)(2k+1)\pi}{4M}\right]$ – импульсный отклик фильтра-анализа

сигнала $x(n)$, обладающий свойствами фильтра нижних частот, или компоненты базисного вектора $h(k)$; 2M – длина кадра.

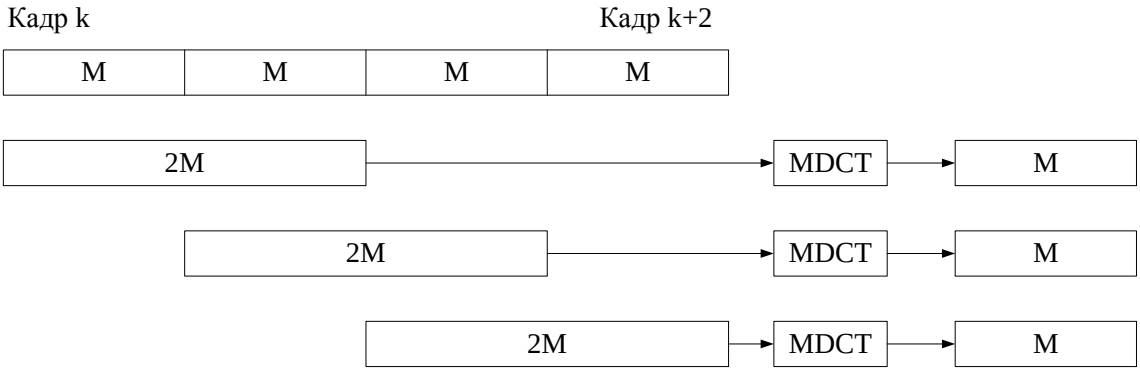
Из (6.26) видно, что прямое МДКП выполняет ряд скалярных произведений между M импульсными откликами фильтра-анализа $h(k)$ и входным сигналом $x(t)$ и вычисляет только M МДКП -коэффициентов для кадра речевого сигнала, состоящего из $2M$ выборок. МДКП основано на использовании пересекающих временных окон анализа сигналов. Наложение во временной области, вводимое с помощью 50% перекрытия окон, устраняется при обратном преобразовании МДКП. Восстановленные выборки $x(n)$ для $0 \le k \le M-1$ вычисляются с обратного МДКП преобразования:

$$x(n) = \sum_{k=0}^{M-1} \left[X(k) h_k(n) + X^P(k)(n+M) \right], \qquad (6.27)$$

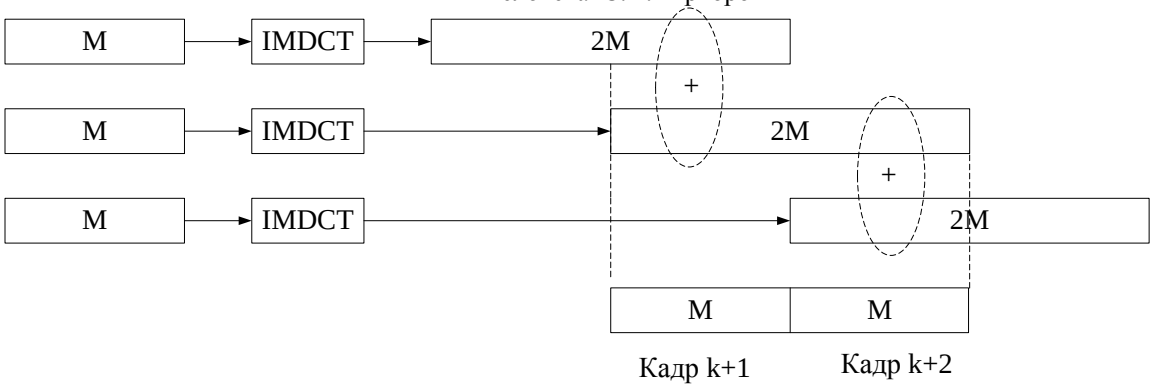
где $X^P(k)$ – предыдущий кадр МДКП коэффициентов.

Из (6.27) видно, что первые M выборок к-го базисного вектора $h(k)$ для $0 \le n \le M-1$ взвешиваются с помощью $к$-го коэффициента текущего кадра $X(k)$. Одновременно вторые M выборок $к$-го базисного вектора $h(k)$ для $M \le n \le 2M-1$ взвешиваются с помощью $к$-го коэффициента предыдущего кадра $X^P(k)$. Затем взвешенные базисные вектора перекрываются и складываются в каждый момент с индексом n.

Перекрывающий анализ и перекрывающий синтез с сложением показан на рис.6.15



а)



б)

Рис.6.15 МДКП преобразование:

а) перекрывающее прямое преобразование (анализ), отображающее 2M выборок в M спектральных коэффициентов, но анализирующий шаг (величина сдвига окна) и время разрешения равно M выборок; б) обратное преобразование, отображающее M спектральных коэффициентов в вектор из 2M выборок, который перекрывается с M выборками и складывается с вектором из 2M выборок предыдущего кадра.

Для МДКП преобразования обобщенное условие полного восстановления может быть сведено к ограничениям Найквиста и линейной фазы на окне:

$$w(2M-1-n) = w(n), \tag{6.28}$$

$$w^2(n) + w^2(n+M) = 1, \tag{6.29}$$

где $0 \le n \le M-1$

Оконные функции. С практической точки зрения ни сигналы, ни реализуемые узкополосные фильтры не имеют строго ограниченной полосы. Следовательно, всегда имеет место некоторое перекрытие соседних частотных полос, что приведет к неизбежным искажениям при восстановлении непрерывного сигнала. Для устранения нежелаемого временного эффекта наложения, называемого *aliasing* (явление, являющееся результатом недостаточной частоты выборки и сопровождаемое появлением ложных частот), окна анализа задаются выражением

$$w(n) = \pm\sqrt{2}\sin\left[\left(n+\frac{1}{2}\right)\frac{\pi}{M}\right]. \tag{6.30}$$

В типичных методах сжатия сигнала, параметры преобразования улучшаются использованием оконной функции w_k (k = 0, ..., 2n-1) которая умножается на x_k и X_k в формулах прямого и обратного преобразования, надлежащим образом избегая разрывов в точках k = 0 и k=2n (границах кадра), делая функцию равномерно сходящейся к нулю в этих точках. В принципе, x и X могут иметь разные оконные функции, и оконная функция может меняться от кадра к кадру (особенно когда комбинируются кадры

разных размеров), но для простоты рассматривается одинаковая оконная функция для одинаковых размеров кадров. Преобразование остается инвертируемым для симметричного окна $w(2M-1-n)=w(n)$, пока w удовлетворяет условию Пирса-Бредли $w^2(n)+w^2(n+M)=1$. Различные оконные функции используются, например $w(n)=\sin\left[\frac{\pi}{2M}\left(n+\frac{1}{2}\right)\right]$ используется в mp3 и MPEG-2 AAC, а $w(n)=\sin\left(\frac{\pi}{2}\sin^2\left[\frac{\pi}{2M}\left(n+\frac{1}{2}\right)\right]\right)$ используется в Vorbis. AC-3 использует окно Кайзера-Бесселя и производное окно (KBD), и MPEG-4 AAC может также использовать KBD окно.

Отметим, что оконные функции, применяемые в МДКП, отличаются от оконных функций, применяемых в других типах разложения сигнала, тем что они должны удовлетворять условию Пирса-Бредли. Одна из причин такого различия заключается в том, что оконные функции для МДКП применяются дважды: в МДКП (разложение) и в ОМДКП (синтез).