

Лекция 6. КОМПРЕССИЯ ЦИФРОВОГО ЗВУКА ИЛИ РЕДУКЦИЯ АУДИОДАНЫХ

При первичном цифровом представлении высококачественных звуковых сигналов суммарная скорость цифрового потока оказывается достаточно большой (при $f_d=48$ кГц и разрешении 16 бит/отсчет) – 768 кбит/с на один сигнал, что требует суммарной пропускной способности канала связи при передаче равной 3,84 Мбит/с.

Пропускная способность слухового аппарата человека оценивается в 100-120 кбит/с, однако сознательно из этого скоростного потока он способен обрабатывать (т.е. воспринимать) лишь около 100 бит/с информации. Поэтому можно говорить о присущей звуковому сигналу избыточности.

Важной проблемой при цифровом представлении ЗС является сокращение имеющейся в них **статистической** и **психофизической** избыточности. Это позволяет уменьшить скорость цифрового потока при кодировании ЗС до предельно возможной величины, при которой шумы, помехи и искажения остаются еще незаметными на слух.

6.1 Статистическая избыточность.

- 1) Сокращение статистической избыточности базируется на учете свойств самих звуковых сигналов. Она обусловлена наличием корреляционной связи между соседними отсчетами звукового сигнала при его дискретизации.

Для ее уменьшения применяют достаточно сложный алгоритм обработки. При их использовании потери информации нет, однако исходный сигнал оказывается представленным в более компактной форме, требующей меньшего количества бит при его передаче.

Наиболее часто для этой цели используют так называемые «ортогональные преобразования».

Оптимальным с этой точки зрения является преобразование Карунена-Лоэва, но его реализация требует значительных вычислительных затрат. Незначительно ему уступает по эффективности модифицированное дискретное косинусное преобразование (**МДКП**). Для него разработаны быстрые вычислительные алгоритмы.

- 2) Уменьшить скорость цифрового потока позволяют также методы кодирования, учитывающие статистику звуковых сигналов, например, вероятности появления уровней разной величины.

Примером такого учета являются коды **Хаффмана**, где наиболее вероятным значениям сигнала приписываются более короткие кодовые слова, а значения отсчетов, вероятность появления которых мала, кодируются кодовыми словами большой длины.

В наиболее эффективных алгоритмах компрессии цифровых аудиоданных кодированию подвергаются не сами отсчеты звукового сигнала, а коэффициенты МДКП, кроме того, при кодировании коэффициентов МДКП часто используются кодовые таблицы Хаффмана.

Но устранение статистической избыточности звуковых сигналов все же не приводит к значительному уменьшению скорости цифрового потока (в конечном итоге удается уменьшить пропускную способность канала связи лишь на 15-25 % по сравнению с ее исходной величиной).

И после устранения статистической избыточности скорость цифрового потока при передаче высококачественных звуковых сигналов и возможности человека по их обработке при слуховом восприятии отличаются на несколько порядков. Это свидетельствует также о существенной психоакустической избыточности первичных звуковых сигналов и, следовательно, о возможности ее уменьшения.

6.2 Психофизическая избыточность.

1) Если известно, какие части ЗС ухо воспринимает, а какие нет (вследствие маскировки) (В определённых случаях один звук может быть скрыт другим звуком. Говорят, что слабый звук маскируется, если он становится неразличимым в присутствии более громкого звука.), то нужно выделить и затем передать по каналу связи лишь те части сигнала, которые ухо способно воспринять, а неслышимые составляющие сигнала можно отбросить.

2) При АЦП сигналы можно квантовать с возможно меньшим разрешением (большим шагом) так, чтобы шумы квантования, изменяясь по величине с изменением уровня сигнала, не становились бы слышимыми.

Учет всех этих свойств слуха позволяет сильно сократить общее число бит (количество информации), требуемое для цифрового представления ЗС.

За счет устранения психофизической избыточности можно уменьшить требуемое для высококачественной передачи (записи) количество цифровых аудиоданных более чем в 10 раз.

Новейшие исследования в области компрессии цифровых аудиоданных привели к появлению новых методов кодирования:

- MUSICAM
- ASPEC
- ATRAC

Разработаны европейские стандарты MPEG-1, MPEG-2, MPEG-4, а также американский стандарт ATSC-AC-3.

В стандартах MPEG имеется несколько профилей и уровней компрессии.

Профили определяют область применения алгоритма, а уровни – степень компрессии цифровых данных (Layer-1, Layer-2 и Layer-3).

Практическое применение методы компактного цифрового представления звуковых сигналов нашли

- 1) в европейской системе цифрового радиовещания DAB;
- 2) в американской ATSC-AC-3 (Dolby AC-3);
- 3) в сетях ISDN.

Достигнуты скорости цифрового потока на выходе кодера с компрессией данных 32...192 кбит/с на канал (уменьшена приблизительно в 25 раз по сравнению с ее исходной величиной 1,5 Мбит/с).

6.3 Психоакустические основы компрессии цифровых аудиоданных.

Учет психоакустических закономерностей слухового восприятия определяет качество систем кодирования с компрессией цифровых данных. Ключевыми моментами при разработке любой психоакустической модели являются:

- 1) вид ортогонального преобразования;
- 2) алгоритм обработки коэффициентов преобразования (БПФ или МДКП) с целью максимально возможного уменьшения числа учитываемых спектральных составляющих;
- 3) выбор аппроксимирующих функций для учета явления маскировки включая проблему суммирования индивидуальных кривых маскировки;
- 4) процедура расчета глобального порога маскировки для спектральной выборки сигнала.

Известны в основном три психоакустические модели: **NMR** (Noise to Mask Ratio – отношение сигнал/маска), **PAQM** – Perceptual Audio Quality Measure, **PERCEVAL** – Perceptual EVALion. Наибольшее распространение пока получила модель **NMR**, в которой учитываются абсолютный порог слышимости и явление маскировки в частотной

области. Это позволяет частично устранить при передаче части звукового сигнала не важные для слухового восприятия – Redundanz und Irrelevanz.

Рассмотрим основные свойства слуха, лежащие в основе NMR-модели.

Абсолютный порог слышимости. Это минимальное значение звукового давления, которое способно еще воспринять человеческое ухо при отсутствии мешающих звуков. Обычно его выражают в дБ по отношению к стандартной величине звукового давления $p_0=2 \cdot 10^{-5}$ Па. Чувствительность слуха к восприятию отдельных спектральных составляющих в очень сильной степени зависит от частоты (рисунок 6.1, нижняя пунктирная кривая). Очевидно, что те спектральные компоненты полезного сигнала, которые лежат ниже абсолютного порога слышимости, кодировать и передавать не следует.

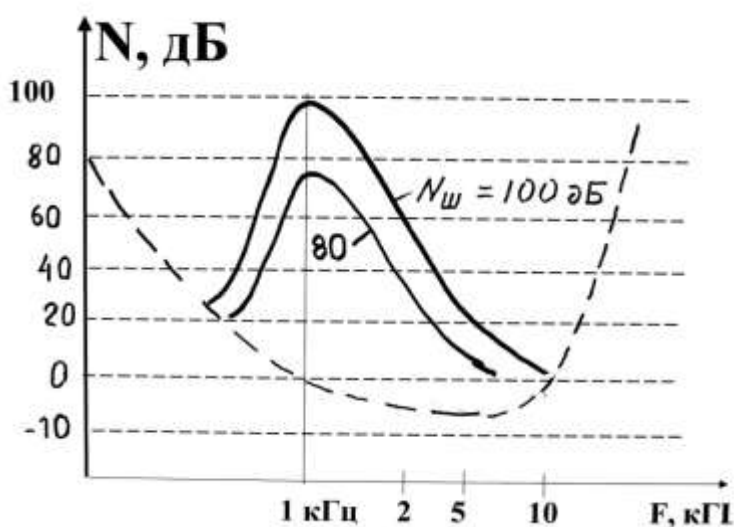


Рисунок 6.1 – Зависимость порога слышимости тона при его маскировке узкополосным шумом

Относительный порог слышимости. Это изменение значения звукового давления, которое способно воспринять человеческое ухо. Известно, что порог слышимости одного сигнала изменяется в присутствии другого сигнала. Это явление носит название маскировки. Этот феномен также поясняет рисунок 6.1 (все кривые, кроме нижней), на котором приведен полученный путем субъективно-статистических экспертиз порог слышимости тона $N_{ПС}$, маскируемого узкополосным белым шумом с уровнем звукового давления $N_{ш}$, равном 20, 40, 60, 80 и 100 дБ. Маскирующий шум имеет полосу частот $\Delta f=160$ Гц, среднюю частоту 1000 Гц. Все эти пять кривых имеют максимум на средней частоте узкополосного шума $F_0=1000$ Гц.

Отметим некоторые общие особенности кривых маскировки:

- с увеличением средней частоты маскирующего шума диапазон частот, где проявляется маскировка, становится тем шире, чем больше уровень маскирующего сигнала $N_{ш}$;
- в точках максимумов кривые маскировки достигают уровней, отличающихся в меньшую сторону от соответствующего уровня маскирующего сигнала на 4 дБ;
- кривые маскировки несимметричны, они имеют крутой спад в сторону нижних частот и пологий спад в сторону высоких частот.

Спектральные компоненты, лежащие ниже относительного порога слышимости слухом, не воспринимаются, поэтому их также можно не передавать на приемную сторону системы при кодировании.

Понятие критической полосы слуха или частотной группы слуха. Если в качестве полезного сигнала выступает тон, а в качестве маскирующего – узкополосный шум, центральная частота которого равна частоте тонального сигнала, и если полоса частот маскирующего шума ΔF расширяется, то оказывается, что при достижении некоторого значения $\Delta F = \Delta F_{\text{чг}}$ величина порога слышимости тона перестанет изменяться. Это значение полосы $\Delta F_{\text{чг}}$ и называют частотной группой слуха. Значение $\Delta F_{\text{чг}}$ зависит от частоты F (рисунок 6.2). На частотах ниже 500 Гц ширина частотной группы постоянна и составляет около 100 Гц.

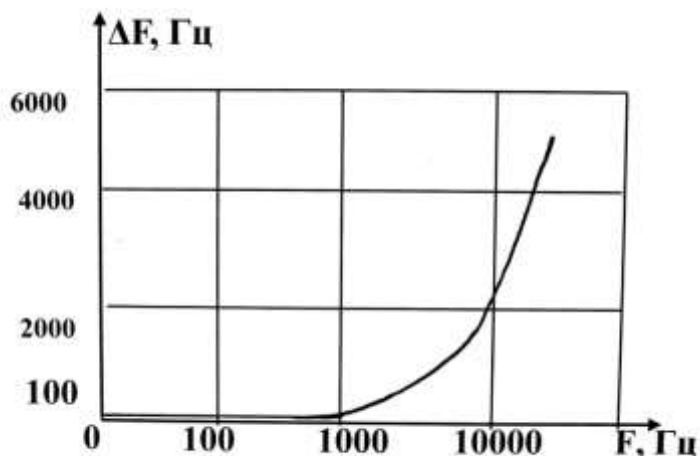


Рисунок 6.2 – Зависимость ширины критической полосы слуха (частотной группы слуха) от средней частоты

В области частот выше 500 Гц она возрастает пропорционально частоте, при этом $\Delta F_{\text{чг}} = 0,2F$. В диапазоне частот от 20 Гц до 16 кГц размещаются 24 частотные группы слуха с шириной от 100 Гц в низкочастотной части и до 4...6 кГц в высокочастотной части спектра. Слух сравнивает полезный сигнал и мешающий шум по интенсивности в пределах критических полос слуха, оценивая порог слышимости.

Маскировка во временной области. Это явление характеризует **динамические** свойства слуха, показывая изменение во времени относительного порога слышимости, когда маскирующий и маскируемый сигналы звучат не одновременно. При этом следует различать явления послемаскировки (изменение порога слышимости после сигнала высокого уровня) и предмаскировки (изменение порога слышимости перед приходом сигнала максимального уровня).

Так, громкое звуковое событие маскирует восприятие более тихих звуков, которые начинаются несколько раньше, т.е. опережают сигнал высокого уровня на интервал времени -30...20 мс (явление предмаскировки) или позже, т.е. запаздывают по времени на 0...120 мс (явление послемаскировки).

При этом имеет значение:

- формы спектров ЗС;
- их уровни;
- их временная последовательность;
- тон это или широкополосный сигнал.

Послемаскировка (или постмаскировка). Явление постмаскировки описывается следующим образом

$$B_M(t) = 1.0 - \frac{1}{1.35} \cdot \arctg \left(\frac{t}{13.2 \cdot T^{0.25}} \right),$$

где $B_M(t)$ – нормированное значение коэффициента послемаскировки; T – длительность маскирующего сигнала, в мс; t – текущее время, мс.

Зависимости $B_M(t)$ показаны на рисунке 6.3. По оси абсцисс отложено текущее время t в мс, по оси ординат – нормированные значения коэффициента постмаскировки. Параметром кривых является длительность T маскирующего сигнала. Видно, что постмаскировка проявляется на интервале времени равном 100...200 мс.

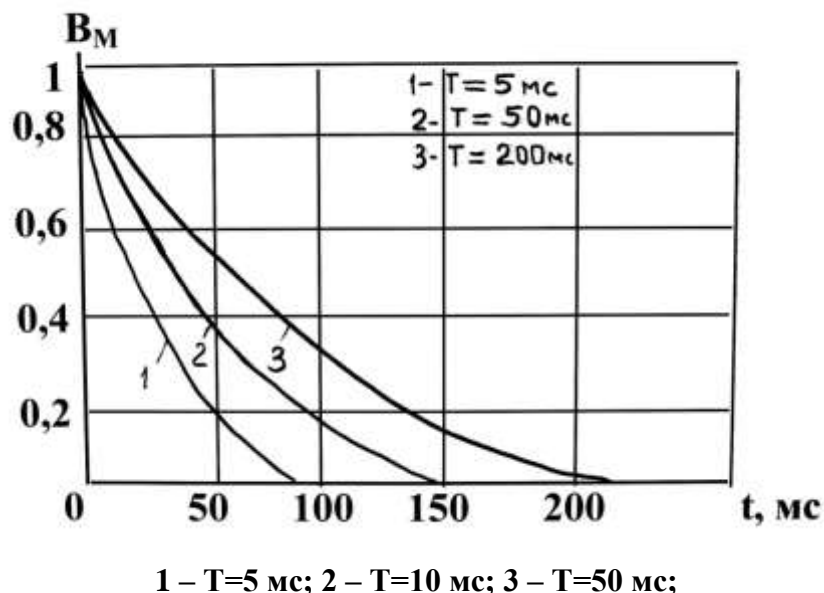


Рисунок 6.3 – Зависимость номинального значения коэффициента постмаскировки от значений текущего времени:

Предмаскировка проявляется на значительно более коротком временном интервале. Он обычно составляет около 10 мс. Длительность предмаскировки в очень сильной степени зависит от особенностей индивидуума. Чаще всего именно по этим двум причинам явление предмаскировки не учитывается.

Глобальный порог маскировки. В психоакустической модели цифровой сигнал разбивается на субполосы. Для каждой субполосы кодирования рассчитывается допустимое значение уровня шумов квантования (так называемое отношение сигнал/маска SMR), при котором они еще маскируются полезным сигналом и не замечаются слухом.

Результат таких вычислений для одного аудиофрейма показан на рисунке 6.4. Верхний рисунок (а) представляет собой выраженную в дБ энергию сигнала выборки в каждой из субполос кодирования. На нижнем рисунке (б) изображены уровень глобального порога маскировки в субполосах психоакустического анализа (кривая вверх) и гистограмма, показывающая максимально-допустимое значение шумов квантования в субполосах кодирования, при котором они еще маскируются полезным сигналом выборки. После этого биты распределяются для субполосных сигналов так, что уровень шумов квантования не превысил бы найденные допустимые значения в каждой из 32 полос. Результат распределения бит для этого же аудиофрейма представлен на рисунке 6.5. Цифры на гистограмме – это число бит, которые следует выделить для кодирования субполосных отсчетов в каждой из полос кодирования, чтобы не был бы превышен глобальный порог маскировки.

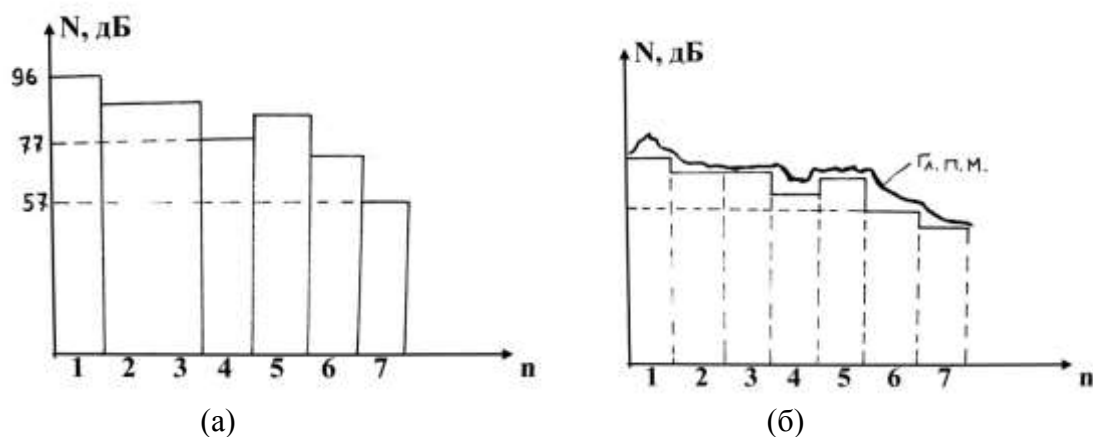


Рисунок 6.4 – Энергия сигнала выборки (а) и глобальный порог маскировки (б) в полосах психоакустического анализа (верхняя кривая) и в полосах кодирования (гистограмма).

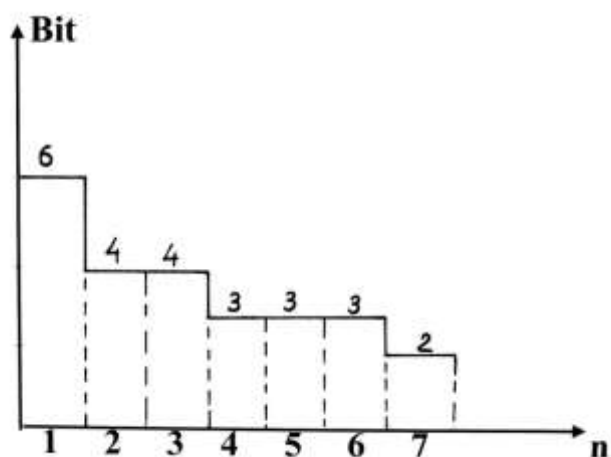


Рисунок 6.5 – Гистограмма распределения бит при кодировании субполосных отсчетов для выборки звукового сигнала в функции от номера полосы кодирования (психоакустическая модель 1 стандарта ISO/IEC 11172-3 или 13818-3, Layer 2)

6.4 Искажения при компрессии цифровых данных. В психоакустических моделях стандартов MPEG не учтены:

- 1) механизм временной маскировки сигналов;
- 2) механизм пространственной демаскировки источников звука, составляющих стереопанораму как по фронту, так и по глубине;
- 3) особенности восприятия реверберационных составляющих стереофонических сигналов.

Именно эти механизмы пространственного слуха играют наиболее важную роль при стереовоспроизведении, они определяют восприятие основных притоков качества стереофонического звучания, таких как пространственное впечатление, прозрачность звучания, естественность и богатство тембров инструментов и голосов, восприятие акустической атмосферы первичного помещения (концертный зал, студия) и т.п.

Предварительные сравнительные прослушивания показали, что компрессия цифровых аудиоданных неизбежно сопровождается следующими искажениями:

- стереопанорама становится более плоской;
- ее пространственная структура нарушается;

- исчезают звуковые планы;
- кажущиеся источники звука смещаются относительно их истинного положения;
- теряются индивидуальные признаки звучания, свойственные однотипным инструментам в группе;
- нарушается реверберационный фон каждого из них, возникает ощущение, что отдельные музыкальные инструменты записаны в разных акустических помещениях;
- теряются индивидуальные признаки первичного помещения записи.

Появление этой совокупности искажений приводит к снижению качества звучания, которое отчетливо отмечается слушателями. Для устранения этих искажений прежде всего необходимо выявить причины их появления, понять их природу и сущность. Только после этого можно будет найти методы для их уменьшения или маскирования – либо путем изменения самой процедуры компрессии цифровых аудиоданных при кодировании, либо путем дополнительной обработки стереофонических сигналов после их декодирования.

6.5 Структура кодера, реализующего алгоритм компрессии аудиоданных

Несмотря на значительное разнообразие алгоритмов компрессии цифровых аудиоданных, структура кодера, реализующего такой алгоритм обработки сигналов, может быть представлена в виде обобщенной схемы, показанной на рис. 6.6.



Рис. 6.6. Обобщенная структурная схема кодера с компрессией цифровых аудиоданных

В блоке временной и частотной сегментации исходный звуковой сигнал разделяется на субполосные составляющие и сегментируется по времени. Длина кодируемой выборки зависит от формы временной функции звукового сигнала. При отсутствии резких выбросов по амплитуде используется так называемая длинная выборка, обеспечивающая высокое разрешение по частоте. В случае же резких изменений амплитуды сигнала длина кодируемой выборки резко уменьшается, что дает более высокое разрешение по времени. Решение об изменении длины кодируемой выборки принимает **блок психоакустического анализа**, вычисляя значение психоакустической энтропии сигнала. После сегментации субполосные сигналы **нормируются, квантуются и кодируются**. В наиболее эффективных алгоритмах компрессии кодированию подвергаются не сами отсчеты выборки ЗС, а соответствующие им коэффициенты МДКП.

Обычно при компрессии цифровых аудиоданных используется энтропийное кодирование, при котором одновременно учитываются как свойства слуха человека, так и

статистические характеристики звукового сигнала. Однако основную роль при этом играют процедуры устранения психоакустической избыточности. Учет закономерностей слухового восприятия звукового сигнала выполняется в **блоке психоакустического анализа**. Здесь по специальной процедуре для каждого субполосного сигнала рассчитывается максимально допустимый уровень искажений (шумов) квантования, при котором они еще маскируются полезным сигналом данной субполосы. **Блок динамического распределения бит** в соответствии с требованиями психоакустической модели для каждой субполосы кодирования выделяет такое минимально возможное их количество, при котором уровень искажений, вызванных квантованием, не превышает порога их слышимости, рассчитанного психоакустической моделью. В современных алгоритмах компрессии используются также специальные процедуры в форме итерационных циклов, позволяющие управлять величиной энергии искажений квантования в субполосах при недостаточном числе доступных для кодирования бит.