

# **Лекции 15 – 16.**

## **Сжатие звуковой информации**

### **План лекции**

- 1. Общие сведения.**
- 2. Структура кодера с компрессией цифровых аудиоданных.**
- 3. Психоакустические модели (ПАМ).**
- 4. Базовые системы кодирования.**

1. Методы сжатия звука основаны на устранении его избыточности. Различают статистическую и психоакустическую избыточность натуральных звуковых сигналов. Сокращение статистической избыточности базируется на учете свойств самих звуковых сигналов, а психоакустической – на учете свойств слухового восприятия.

**Статистическая избыточность** обусловлена наличием корреляционной связи между соседними отсчетами временной функции звукового сигнала (ЗС) при его дискретизации. Для ее уменьшения применяют достаточно сложные алгоритмы обработки. При их использовании потери информации нет, однако исходный сигнал оказывается представленным в более компактной

форме, что требует меньшего количества бит при его кодировании. Однако даже при использовании достаточно сложных процедур обработки устранение статистической избыточности звуковых сигналов позволяет увеличить требуемую пропускную способность канала связи лишь на 15...25% по сравнению с ее исходной величиной, что нельзя считать революционным достижением.

После устранения статистической избыточности скорость цифрового потока при передаче высококачественных ЗС и возможности человека по их обработке отличаются, по крайней мере, на несколько порядков.

Это свидетельствует также о существенной **психоакустической избыточности** первичных цифровых ЗС и, следовательно, о возможности ее уменьшения. Наиболее перспективными с этой точки зрения оказались методы, учитывающие такие свойства слуха, как маскировка. Если известно, какие части звукового сигнала ухо воспринимает, а какие нет вследствие маскировки, то можно

выделить и затем передать по каналу связи лишь те части сигнала, которые ухо способно воспринять, а неслышимые – можно просто отбросить. Кроме того, сигналы можно квантовать с возможно меньшим разрешением по уровню так, чтобы искажения квантования, изменяясь по величине с изменением уровня самого сигнала, еще оставались

бы неслышимыми - маскировались бы  
исходным сигналом. Однако после  
устранения психоакустической избыточности  
точное восстановление формы временной  
функции ЗС при декодировании оказывается  
уже невозможным.

Две важные для практики особенности:

- Если компрессия цифровых аудиосигналов уже использовалась в канале связи, то ее повторное применение ведет к появлению существенных искажений, т.е. важно знать «историю» цифрового сигнала и какие методы кодирования уже использовались.

- Традиционные методы оценки качества (например, на тональных сигналах) для кодеков с компрессией цифровых аудиоданных не пригодны, поскольку тестирование проводится на реальных звуковых сигналах.

Работы по анализу качества и оценке эффективности алгоритмов компрессии цифровых аудиоданных с целью их последующей стандартизации начались в 1988 году, когда была образована международная экспертная группа MPEG (Moving Pictures Experts Group).

Итогом работы этой группы на первом этапе явилось принятие в ноябре 1992 года международного, стандарта MPEG-1 ISO/IEC 11172-3 (цифра 3 после номера стандарта относится к кодированию звуковых сигналов).

К настоящему времени достаточное распространение получили еще нескольких стандартов MPEG, таких, как MPEG-2 ISO/IEC 13818-3, 13818-7 и MPEG-4 ISO/IEC 14496-3. В отличие от этого в США, в качестве альтернативны стандартам MPEG, был разработан стандарт Dolby AC-3.

Несколько позже четко сформировались две разные платформы цифровых технологий для радиовещания и телевидения – это DAB (Digital Audio Broadcasting), DRM (Digital Radio Mondiale), DVB (с наземной DVB-T, кабельной DVB-C, спутниковой DVB-S разновидностями) и ATSC (Dolby AC-3).

Первая из них (DAB, DRM) продвигается Европой, ATSC – США. Отличаются эти платформы, прежде всего, выбранным алгоритмом компрессии цифровых аудиоданных, видом цифровой модуляции и процедурой помехоустойчивого кодирования ЗС.

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

# 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 временных характеристик звукового сигнала.

При отсутствии резких выбросов по амплитуде используется так называемая длинная выборка. В случае же резких изменений амплитуды сигнала длина кодируемой выборки существенно 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.

Энтропийное кодирование (при  
компрессии цифровых аудиосигналов)  
учитывает и свойства слуха, и статистические  
характеристики звукового сигнала.

Но основную роль играет  
психоакустический анализ закономерностей  
слухового восприятия.

3. Качество психоакустического анализа определяется психоакустической моделью (ПАМ).

Наиболее известны три модели ПАМ:

- NMR (Noise to Mask Radio);
- RAQM (Perceptual Audio Quality Measure);
- PERCEVAL (PERCeptual EVAListion).

Наиболее распространена NMR – модель, применяющаяся в стандартах MPEG-1 и MPEG-2 для уровней (слоев) компрессии Level 1 и Level 2.

Модель NMR использует следующие свойства слуха:

- **Абсолютный порог слышимости.**
- **Критические полосы слуха** (частотные группы, на которые человек разделяет звуковой сигнал при его восприятии), у которых даже есть своя единица измерения для высоты тона – барк.

- **Относительный порог слышимости и маскировка в частотной области.** При одновременном воздействии на слух двух сигналов один на фоне другого может быть не слышен – это *маскировка*, а *относительный порог слышимости* – это порог слышимости одного сигнала в присутствии другого с учетом маскировки по частоте

- **Маскировка во временной области** – характеризует динамические свойства слуха, показывая изменение во времени относительного порога слышимости, когда маскирующий и маскируемый сигналы звучат не одновременно.

При этом различают *послемаскировку* (изменение порога слышимости после сигнала высокого уровня) и *предмаскировку* (изменение порога слышимости перед приходом сигнала высокого уровня).

Такой вид маскировки, когда звуки не перекрываются во времени, называется *временной маскировкой*.

Послемаскировка проявляется на интервале времени 100...200 мс после окончания маскирующего сигнала, а предмаскировка – около 10 мс, что определяется особенностями конкретного человека. По этой причине временная маскировка при цифровом кодировании практически не используется.

Основные процедуры вычислений, которые выполняются на базе психоакустического анализа, реализованного на основе NMR – модели по принципу аддитивного (взаимонезависимого) действия на орган слуха спектральных компонент, если они воздействуют одновременно.

На вход блока психоакустического анализа кодера (слайд 17) подается первичный ИКМ сигнал

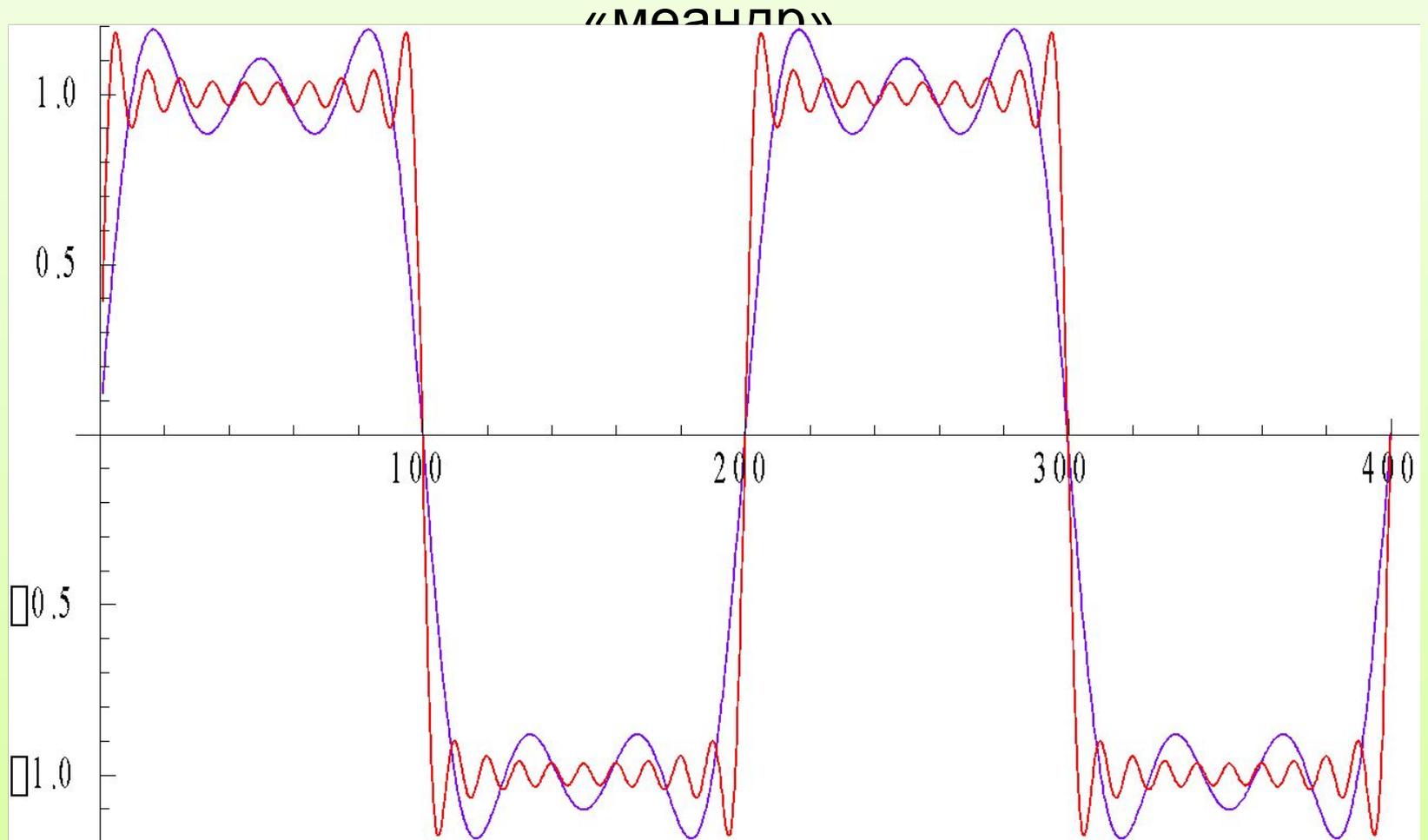
со скоростью  $48 \cdot 16 = 768$  Кбит/с. Выполняются следующие процедуры:

**Процедура 1. Расчет энергетического спектра выборки входного ЗС и его нормирование.**

Пример: пусть длина выборки БПФ  $N=512$  (Layer 1) или 1024 отсчета (Layer 2). Обозначим  $n$  - номер отсчета сигнала в выборке;  $k$  - индекс коэффициента БПФ.

На выходе блока БПФ имеем линейчатый спектр  $X(k)$  в дБ, с разрешением по частоте  $\Delta F = f_d/N$ . При  $f_d = 48$  кГц и  $N = 1024$  получим  $\Delta F = 46,875$  Гц. БПФ выполняется с оконной функцией Hanning для подавления эффекта Гиббса.

# Эффект Гиббса для прямоугольного сигнала типа «меандр»



Вычисленный спектр нормируется, максимальной спектральной компоненте присваивается уровень 92 дБ.

**Процедура 2. Вычисление энергии сигнала выборки в субполосах кодирования.**

**Процедура 3. Вычисление локальных максимумов энергетического спектра сигнала выборки.** Алгоритм здесь простой: спектральная компонента  $X(k)$  будет локальным максимумом,

Если она больше предшествующей  $X(k-1)$ , но не менее следующей  $X(k+1)$ .

**Процедура 4. Формирование списка тональных компонент.** В этом случае исследуется область частот около каждого локального максимума и соответствующая спектральная составляющая включается в список тональных компонент  $\{X(k)\}$ , если в этой области она

превышает любую компоненту (кроме двух соседних, чтобы учесть при расчете уровня их энергии) не менее чем на 7 дБ.

**Процедура 5. Формирование списка нетональных (шумоподобных) компонент** осуществляется после формирования списка тональных компонент. Для этого из исходного спектра сигнала выборки исключаются тональные и

соседние компоненты, учтенные ранее. Данная процедура необходима, чтобы учесть соответствующие коэффициенты маскировки.

**Процедура 6. Прореживание спектра тональных и нетональных компонент осуществляется с целью маскировки вне критической полосы слуха, которая одинакова и для тональных и для нетональных компонент.**

После прореживания формируется новая сетка спектральных компонент: в первых трех субполосах (0...2250 Гц) учитываются все спектральные компоненты, в следующих трех субполосах (2250...4500 Гц) – каждая вторая, в последующих трех субполосах (4500...6750 Гц) – каждая четвертая и в оставшихся 20 субполосах – лишь каждая восьмая спектральная компонента.

Таким образом, если верхняя частота ЗС 22500Гц, то после такого прореживания получается спектр из 126 спектральных компонент (исходный спектр имел 512 составляющих).

**Процедура 7. Расчет коэффициентов маскировки.**

**Процедура 8. Расчет порогов маскировки.**

**Процедура 9. Вычисление кривой глобального порога маскировки.** Здесь формируется глобальный порог маскировки для каждой субполосы и определяется допустимое значение уровня шумов для каждого квантования, в частности, строится гистограмма распределения бит при кодировании субполосных отсчетов.

**4.1.** Звуковая часть стандарта MPEG-1 (ISO/IEC 11172-3) включает в себя три алгоритма различных уровней сложности: Layer (уровень) I, Layer II и Layer III. Общая структура процесса кодирования одинакова для всех уровней, но они различаются по целевому использованию и внутренним механизмам. Для каждого уровня определен свой цифровой поток, то есть общая

ширина потока и свой алгоритм декодирования. Уровни имеют различия в обеспечиваемом коэффициенте сжатия и качестве звучания получаемых потоков.

MPEG-1 предназначен для кодирования сигналов, оцифрованных с частотой дискретизации 32, 44.1 и 48 кГц.

Стандарт MPEG-1 нормирует для всех трех уровней следующие номиналы скоростей цифрового потока: 32, 48, 56, 64, 96, 112, 192, 256, 384 и 448 кбит/с, число уровней квантования входного сигнала – от 16 до 24.

Стандартным входным сигналом для кодера MPEG-1 принят цифровой сигнал AES/EBU (двухканальный цифровой звуковой сигнал с разрядностью квантования 20 ... 24 бита на отсчет). Предусматриваются следующие режимы работы звукового кодера: одиночный канал (моно), двойной канал (стерео или два моноканала) и

joint stereo (сигнал с частичным разделением правого и левого каналов).

Важнейшим свойством MPEG-1 является полная обратная совместимость всех трех уровней. Это означает, что каждый декодер может декодировать сигналы не только своего, но и нижележащих уровней.

В основу алгоритма Уровня I положен формат DCC (Digital Compact Cassette), разработанный компанией Philips для записи на компакт-кассеты. Кодирование первого уровня применяется там, где не очень важна степень компрессии и решающими факторами являются сложность и стоимость кодера и декодера.

Кодер Уровня I обеспечивает  
высококачественный звук при скорости  
цифрового потока 384 кбит/с на  
стереопрограмму.

Уровень II требует более сложного кодера и  
несколько более сложного декодера, но  
обеспечивает лучшее сжатие –

«прозрачность» канала достигается уже при скорости 256 кбит/с. Он допускает до 8 кодирований/декодирований без заметного ухудшения качества звука. В основу алгоритма Уровня II положен популярный в Европе формат MUSICAM.

Самый сложный Уровень III включает все основные инструменты сжатия: полосное кодирование, дополнительное ДКП, энтропийное кодирование, усовершенствованную ПАМ. За счет усложнения кодера и декодера он обеспечивает высокую степень компрессии – считается, что «прозрачный» канал формируется уже на скорости

128 кбит/с, хотя высококачественная передача возможна и на более низких скоростях.

В стандарте рекомендованы две психоакустические модели: более простая Модель 1 и более сложная, но и более высококачественная Модель 2. Они отличаются алгоритмом обработки отсчетов. Обе модели могут использоваться для всех трех уровней,

но Модель 2 имеет специальную  
модификацию для Уровня III.

MPEG-1 оказался первым международным  
стандартом цифрового сжатия звуковых  
сигналов и это обусловило его широкое  
применение во многих областях:

вещании, звукозаписи, связи и мультимедийных приложениях. Наиболее широко используется Уровень II, он вошел составной частью в европейские стандарты спутникового, кабельного и наземного цифрового ТВ вещания, в стандарты звукового вещания, записи на DVD,

## Рекомендации МСЭ BS.1115 и J.52.

Уровень III (его еще называют MP-3) нашел широкое применение в цифровых сетях с интегральным обслуживанием (ISDN) и в сети Интернет. Подавляющее большинство музыкальных файлов в сети записаны именно в этом стандарте.

4.2. MPEG-2 это расширение MPEG-1 в сторону многоканального звука.

MPEG-2 учитывает различия режима передачи многоканального звука, в том числе пятиканальный формат, семиканальный звук

с двумя дополнительными  
громкоговорителями, применяемыми в  
кинотеатрах с очень широким экраном,  
расширения этих форматов с  
низкочастотным каналом.

4.3. При всем множестве новаторских подходов MPEG-4 звуковые разделы стандарта – возможно, наиболее интересная и революционная его часть. Объектный подход к изображениям – новое для телевидения, но в ряде систем анимации он применялся и ранее.

По поводу звуковых качество стандарта (т. н. объектного звука), то системы, сопоставимой с MPEG-4 по комплексности подхода, спектру примененных технологий и диапазону применений, просто нет.

Принципиальным отличием MPEG-7 является то, что он разрабатывался совсем не для установления каких-либо правил сжатия аудио- и видеоданных или типизацию и характеристику данных какого-то конкретно рода.

4.4. Стандарт MPEG-7 предусмотрен как описательный, предназначенный для регламентации характеристик данных мультимедиа любого типа, вплоть до аналоговых, и записанных в разных форматах (например, с разным пространственным и временным разрешением кадра).