Сжатие Аудиоданных

Общие принципы и устройство MP3

Дмитрий Ватолин

Московский Государственный Университет
CS MSU Graphics&Media Lab

Version 2.1

Благодарности

- Автор выражает глубокую признательность Алексею Лукину и Александру Жиркову (Graphics&Media Lab) за предоставленные слайды лекций

Сжатие аудио

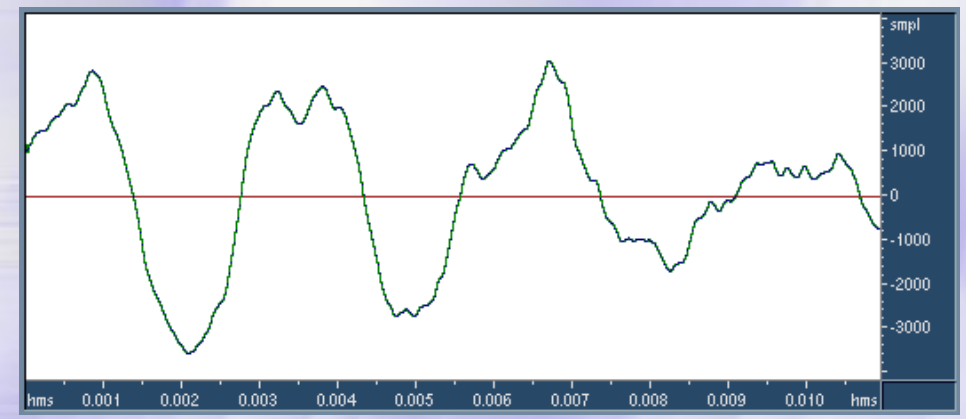
GRAPHICS & MEDIA LAB
VIDEO GROUP

- Общие понятия и принципы сжатия с потерями, психоаккустика.
- Устройство алгоритма MP3
- Гибридные методы сжатия
- Речевые кодеки

Сигналы

- Сигнал – скалярная функция от одного или нескольких аргументов.

примеры сигналов:



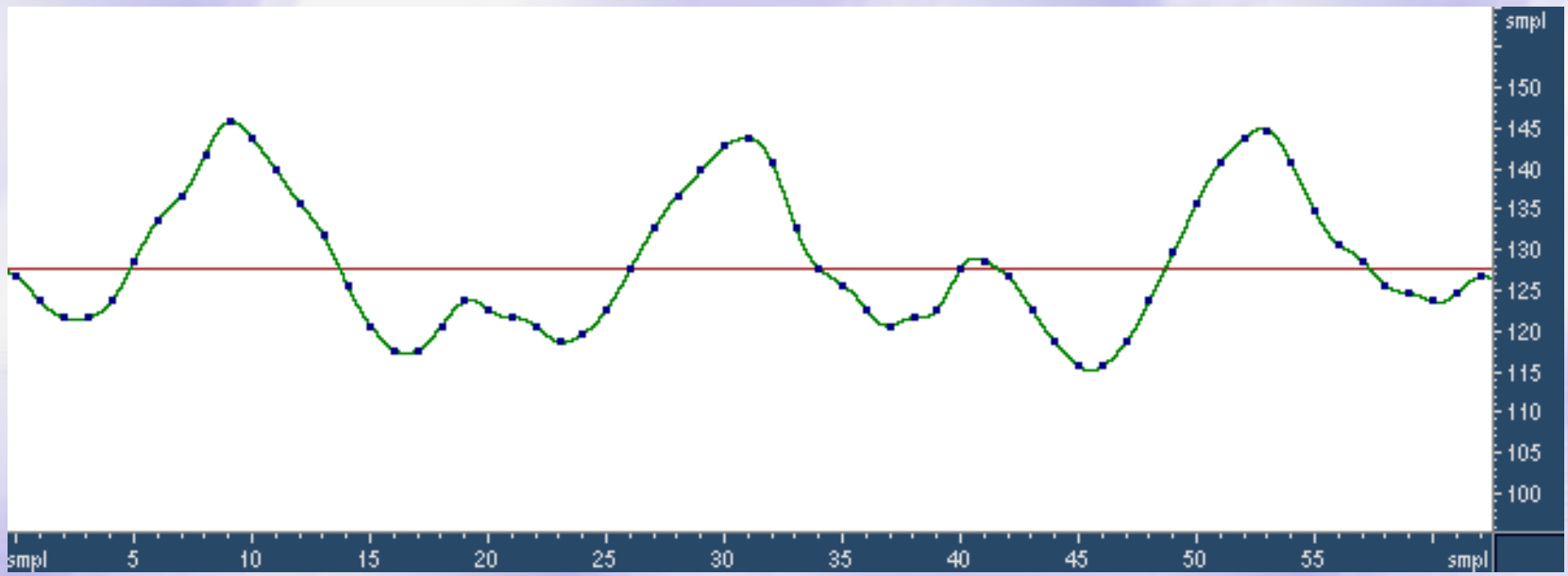
s(t) – звук

f(x,y) – изображение

Оцифровка сигналов

GRAPHICS & MEDIA LAB
VIDEO GROUP

1. Дискретизация по времени
2. Квантование по амплитуде



Оцифровка сигналов

GRAPHICS & MEDIA LAB
VIDEO GROUP

- При каких условиях по цифровому сигналу можно точно восстановить исходный аналоговый?
- Предположим, что значения амплитуд в цифровом сигнале представлены точно.
- Введем понятие спектра аналогового сигнала:

(разложение на синусоиды с различными частотами)

$$x(t) = \int_{-\infty}^{+\infty} X(\nu) \cdot e^{2\pi i \nu t} d\nu \qquad X(\nu) = \int_{-\infty}^{+\infty} x(t) \cdot e^{-2\pi i \nu t} dt$$

x(t) – исходный сигнал
X(*ν*) – спектр, т.е. коэффициенты при гармониках с частотой *ν*

Теорема Котельникова

GRAPHICS & MEDIA LAB
VIDEO GROUP

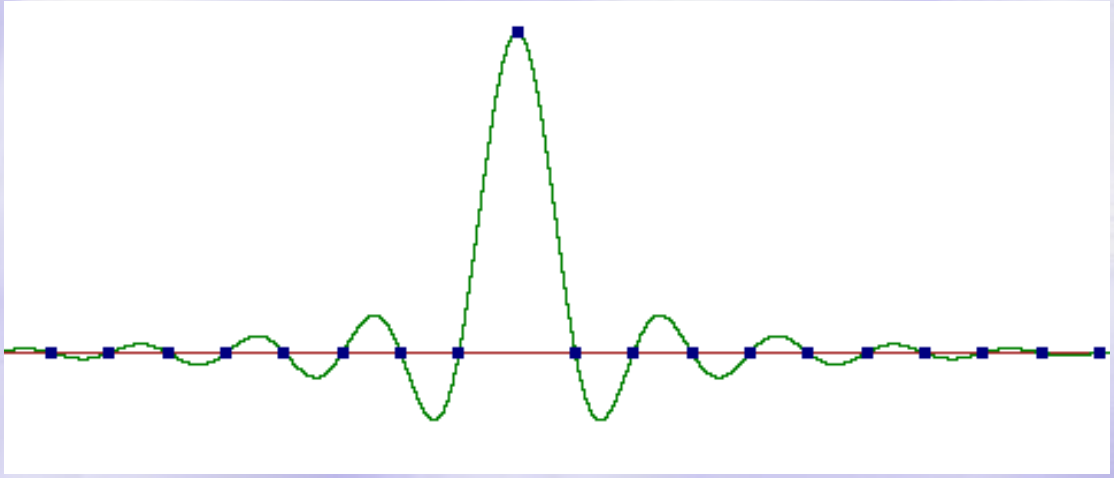
- Пусть:
  1. спектр сигнала x(t) не содержит частот выше F, т.е. X(ν)=0 за пределами отрезка [-F, F]
  2. дискретизация сигнала x(t) производится с частотой F_s, т.е. в моменты времени nT, здесь $T= F_s^{-1}$
  3. $F_s \geq 2F$
- Тогда исходный аналоговый сигнал x(t) можно точно восстановить из его цифровых отсчетов x(nT), пользуясь интерполяционной формулой

$$x(t) = \sum_{n=-\infty}^{+\infty} x(nT) \cdot Sinc(t - nT) \qquad Sinc(t) = \frac{\sin \pi F_s t}{\pi F_s t}$$

Теорема Котельникова

- Как выглядят интерполирующие функции?

$$x(t) = \sum_{n=-\infty}^{+\infty} x(nT) \cdot Sinc(t - nT) \qquad Sinc(t) = \frac{\sin \pi F_s t}{\pi F_s t}$$



Бесконечно затухающие колебания

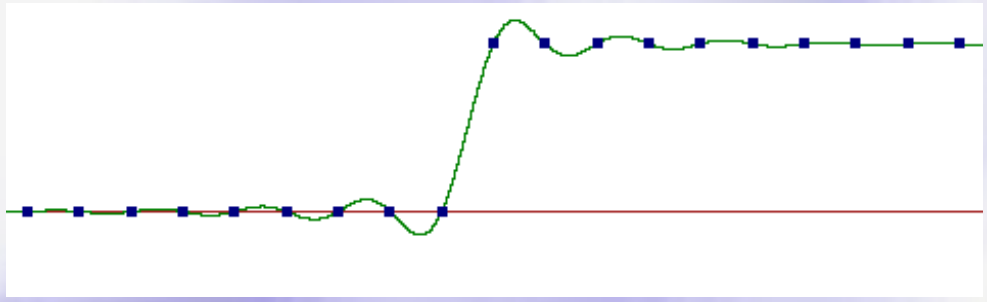
Теорема Котельникова

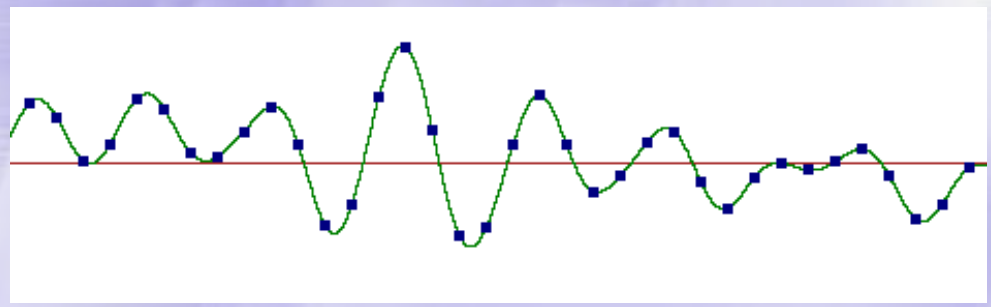
GRAPHICS & MEDIA LAB
VIDEO GROUP

- Реконструкция аналоговых сигналов:

$$x(t) = \sum_{n=-\infty}^{+\infty} x(nT) \cdot Sinc(t - nT)$$

sinc-интерполяция

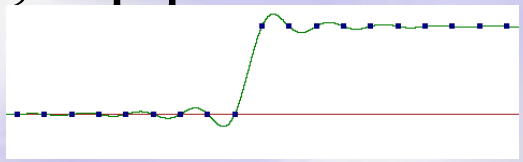




Теорема Котельникова

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Применимость sinc-интерполяции для изображений, эффект Гиббса



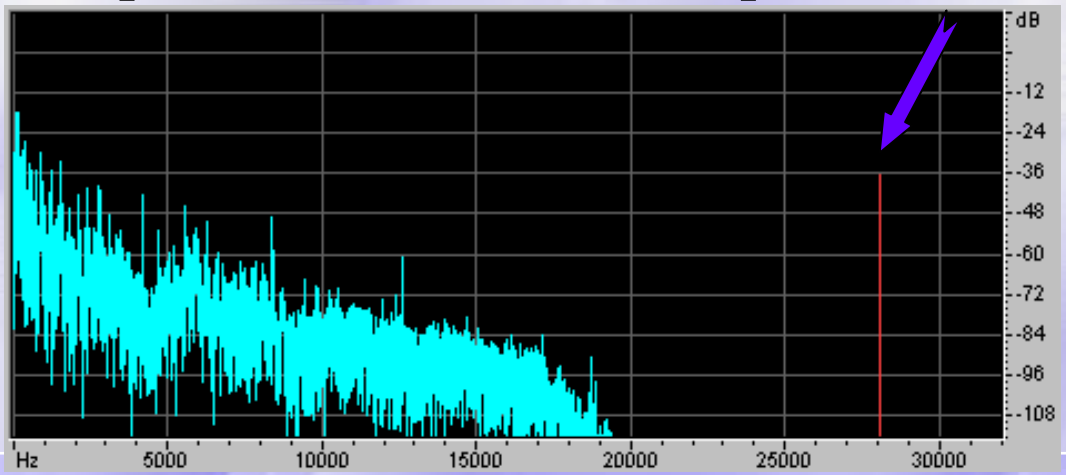
Цифровые отсчеты

sinc-интерполяция

другая интерполяция

Алиасинг

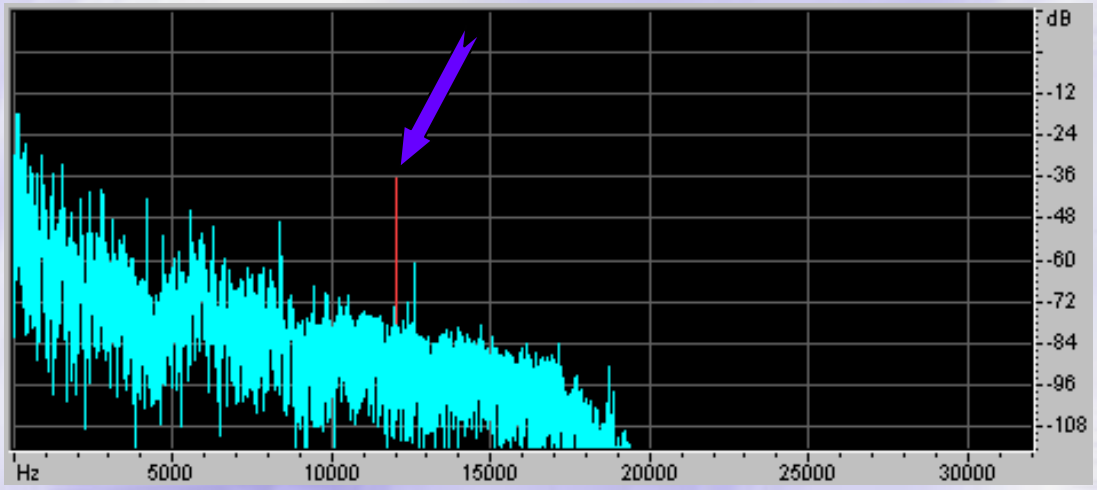
- Что будет, если условия теоремы Котельникова не выполнены?
- Пусть звук не содержит частот выше 20 кГц. Тогда, по теореме Котельникова, можно выбрать частоту дискретизации 40 кГц.
- Пусть в звуке появилась помеха с частотой 28 кГц. Условия теоремы Котельникова перестали выполняться.



Алиасинг

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Проведем дискретизацию с частотой 40 кГц, а затем – восстановим аналоговый сигнал sinc-интерполяцией.



- Помеха отразилась от половины частоты дискретизации в нижнюю часть спектра и наложилась на звук. Помеха переместилась в слышимый диапазон. Алиасинг.

Алиасинг

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Как избежать алиасинга?
- Применить перед оцифровкой анти-алиасинговый фильтр:
  - Он подавит все помехи выше половины частоты дискретизации (выше 20 кГц) и пропустит весь сигнал ниже 20 кГц.
  - После этого условия теоремы Котельникова будут выполняться и алиасинга не возникнет.
  - Следовательно, по цифровому сигналу можно будет восстановить исходный аналоговый сигнал.

Преобразование Фурье

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Зачем раскладывать сигналы на синусоиды?
  - Анализ линейных систем
  - Слух и синусоиды
  - Хорошо разработана теория и практика
- Дискретное преобразование Фурье (ДПФ)
- Ряд Фурье

$$x[n] = \sum_{k=0}^{N/2} C_k \cos \frac{2\pi k(n+\varphi_k)}{N}$$

- Частоты и амплитуды

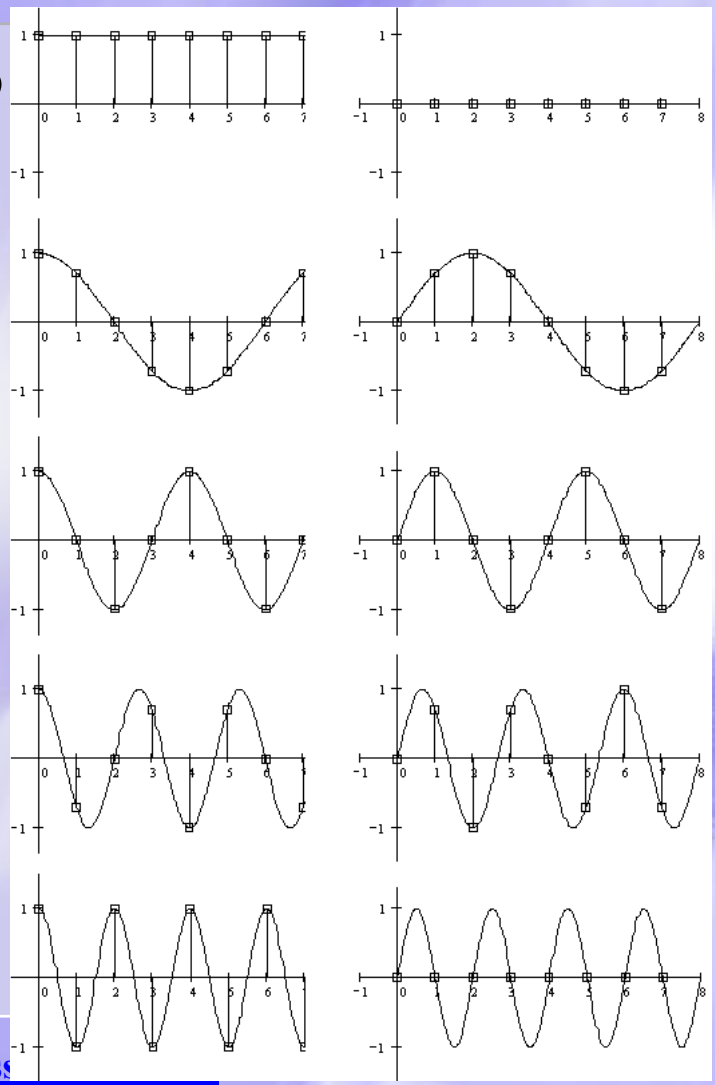
$$x[n] = \sum_{k=0}^{N/2} A_k \cos \frac{2\pi kn}{N} + \sum_{k=0}^{N/2} B_k \sin \frac{2\pi kn}{N}$$

- Прямое и обратное преобразования Фурье

Преобразование Фурье

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Базисные функции дискретного преобразования Фурье для сигнала длины N = 8.
- Имеем N/2 + 1 = 5 различных базисных частот.
- Имеем N+2 базисные функции, 2 из которых тождественно равны нулю.
- Количество информации не изменяется: N чисел



Преобразование Фурье

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Базисные функции образуют N-мерный ортогональный базис в пространстве N-мерных векторов исходных сигналов.
- Следовательно, разложение обратимо, т.е. по коэффициентам разложения (A_k, B_k) можно точно восстановить исходный дискретный сигнал.
- Обратное преобразование Фурье – вычисление суммы конечного ряда Фурье (сложить N штук N-точечных синусоид со своими коэффициентами).

Преобразование Фурье

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Прямое преобразование Фурье – вычисление скалярных произведений сигнала на базисные функции:

$$A_k = \frac{2}{N}\sum_{i=0}^{N-1} x[i]\cos\frac{2\pi k i}{N} \qquad k = 1,\ldots,\frac{N}{2}-1$$

$$A_k = \frac{1}{N}\sum_{i=0}^{N-1} x[i]\cos\frac{2\pi k i}{N} \qquad k = 0,\frac{N}{2}$$

$$B_k = \frac{2}{N}\sum_{i=0}^{N-1} x[i]\sin\frac{2\pi k i}{N} \qquad k = 0,\ldots,\frac{N}{2}$$

- Для вычисления всех коэффициентов по этому алгоритму требуется примерно N^2 умножений: очень много при больших длинах сигнала N.

Преобразование Фурье

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Быстрое преобразование Фурье (БПФ, FFT) – ускоренный алгоритм вычисления ДПФ
  - Основан на периодичности базисных функций (много одинаковых множителей)
  - Математически точен (ошибки округления даже меньше, т.к. меньше число операций)
  - Число умножений порядка $N \cdot \log_2 N$, намного меньше, чем N^2
  - Ограничение: большинство реализаций FFT принимают только массивы длиной $N = 2^m$

Существует и обратное БПФ (IFFT) – такой же быстрый алгоритм вычисления обратного ДПФ.

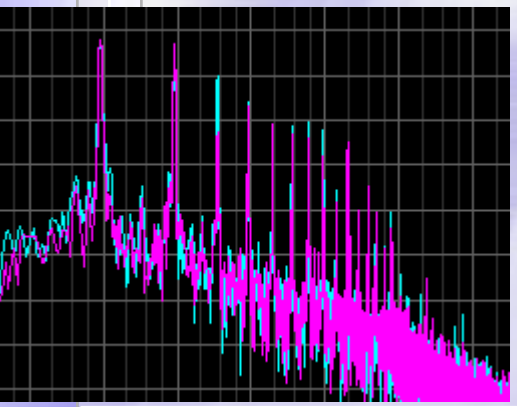
Спектральный анализ

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Отображение спектра звука: спектрограмма
  - Спектрограмма – график зависимости амплитуды от частоты
  - Низкие частоты – слева, высокие – справа
  - Часто применяется логарифмический масштаб частот и амплитуд: “log-log-спектрограмма”
  - Временное и частотное разрешение спектрограммы

Децибелы: $D = 20 \lg \frac{A_1}{A_0}$

A_1 – амплитуда измеряемого сигнала,
A_0 – амплитуда сигнала, принятого за начало отсчета (0 дБ)



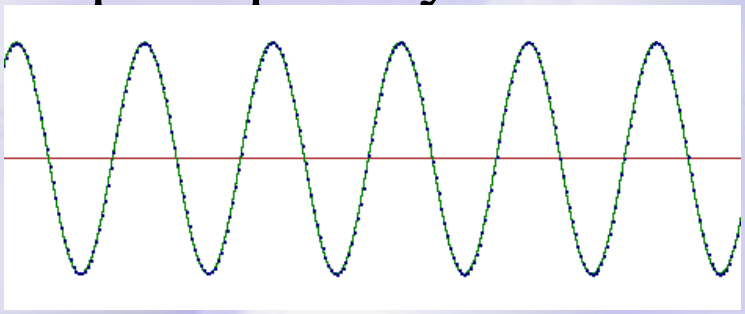
Разница на 6 дБ – разница по амплитуде в 2 раза,
разница на 12 дБ – разница по амплитуде в 4 раза.

Часто за 0 дБ принимается либо самый тихий слышимый звук, либо самый громкий звук, который может воспроизвести аудио-устройство.

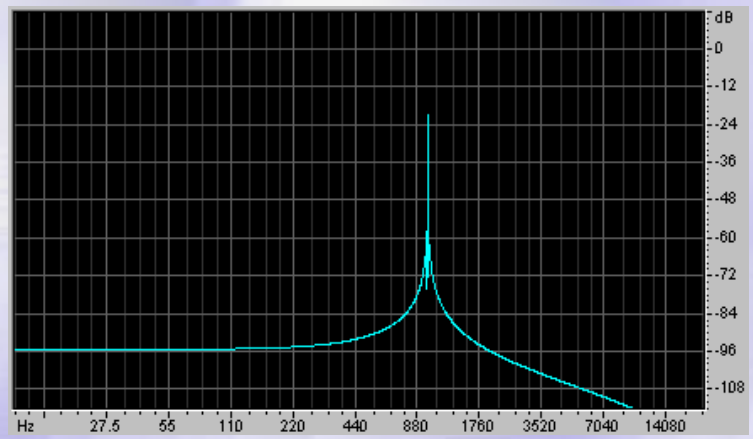
Спектральный анализ

GRAPHICS & MEDIA LAB
VIDEO GROUP

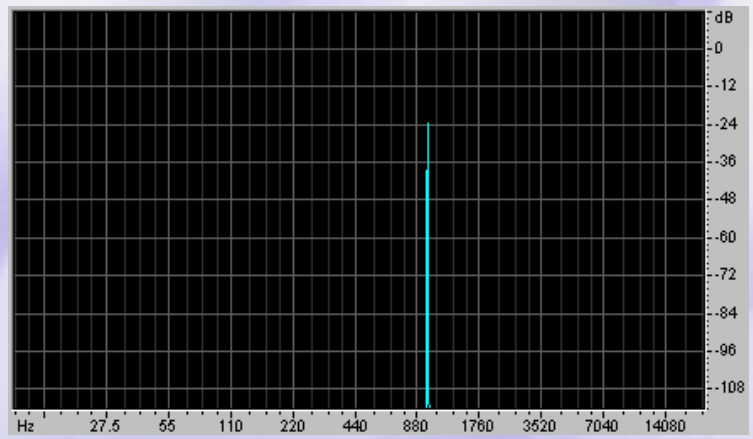
- Примеры звуков и их спектров



Исходная волна – синусоида



Спектр с одним весовым окном

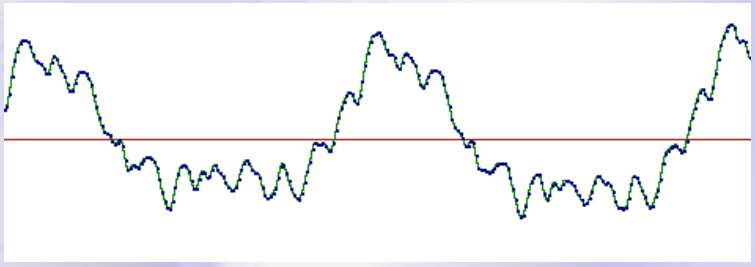


Спектр с другим весовым окном

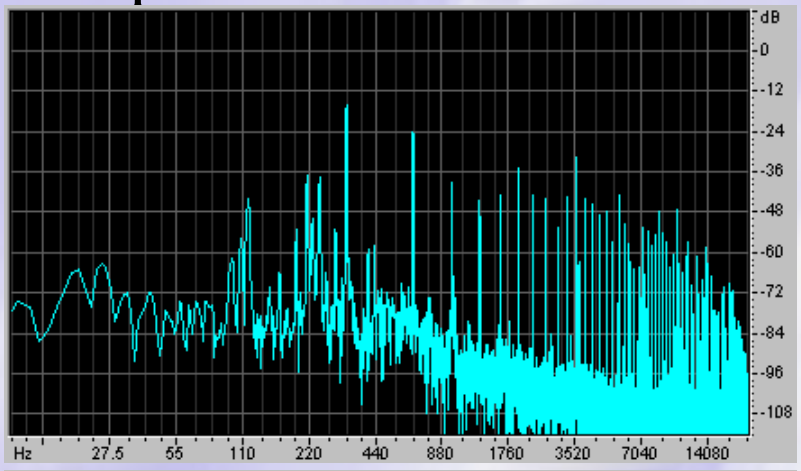
Спектральный анализ

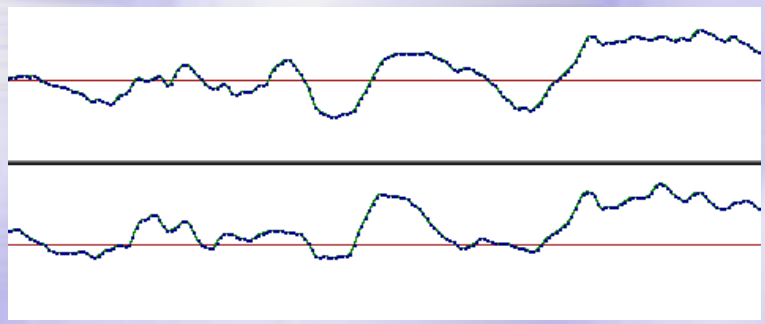
GRAPHICS & MEDIA LAB
VIDEO GROUP

- Примеры звуков и их спектров

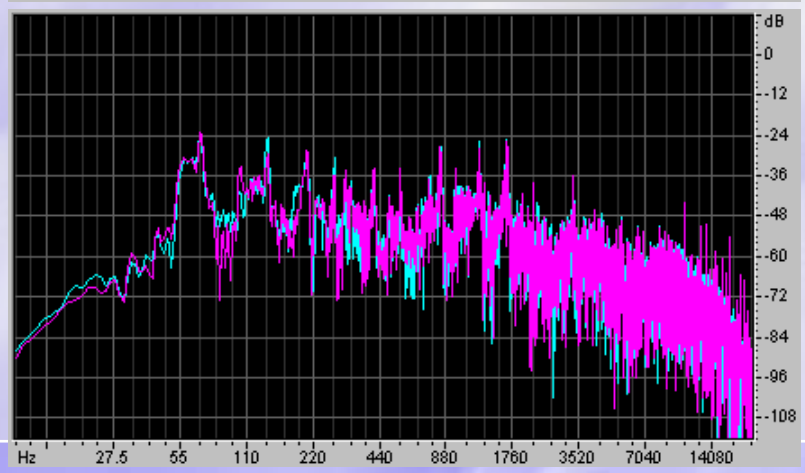


Нота на гитаре





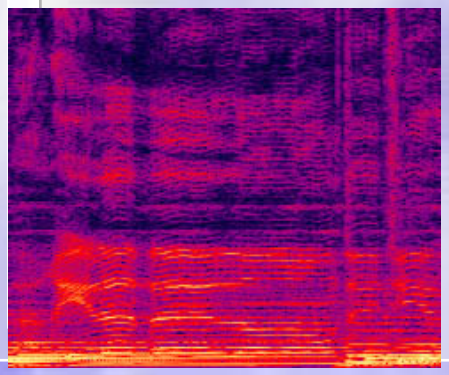
Песня (стерео запись)



Спектральный анализ

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Отображение спектра звука: сонограмма
  - Сонограмма – график зависимости амплитуды от частоты и от времени
  - Низкие частоты – снизу, высокие – сверху
  - Время идет справа налево
  - Амплитуда – яркость или цвет
  - Частотное и временное разрешение
  - Short Time Fourier Transform (STFT)

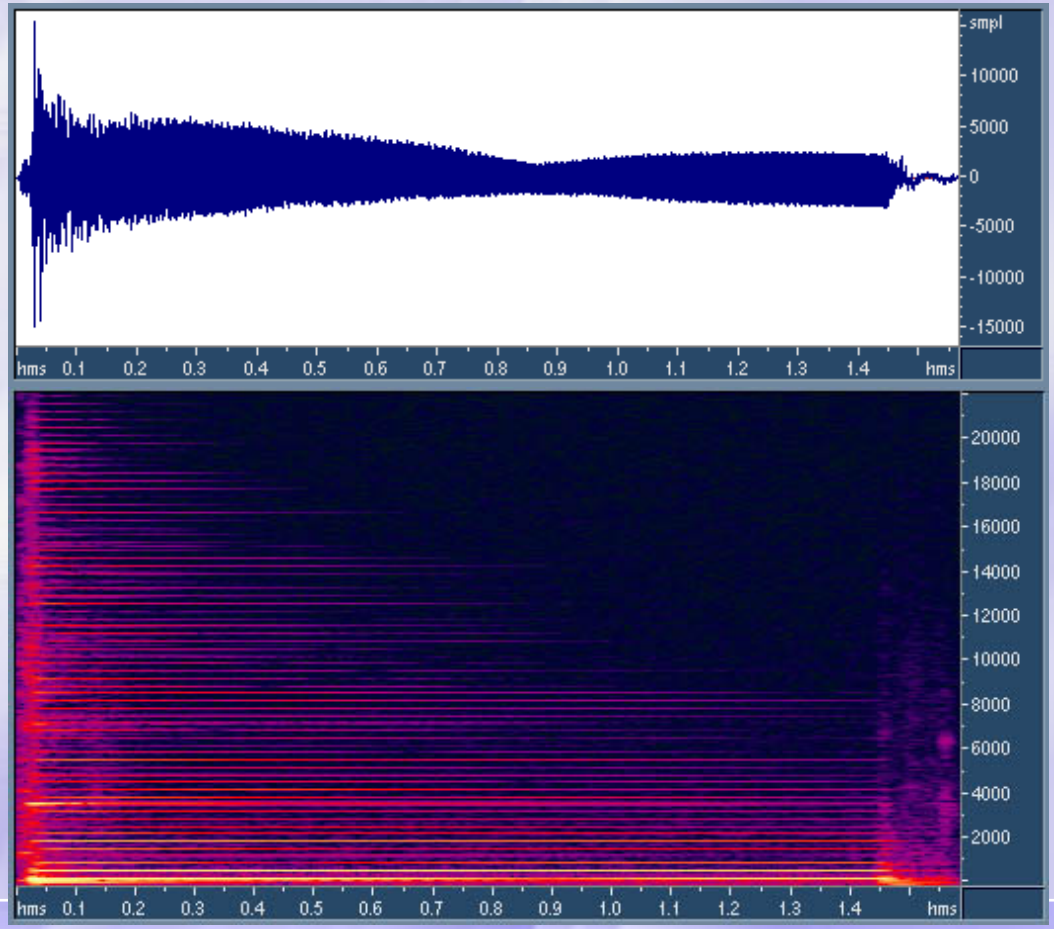


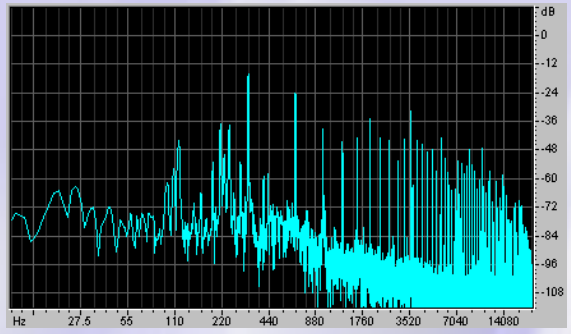
Показывает изменение спектра во времени

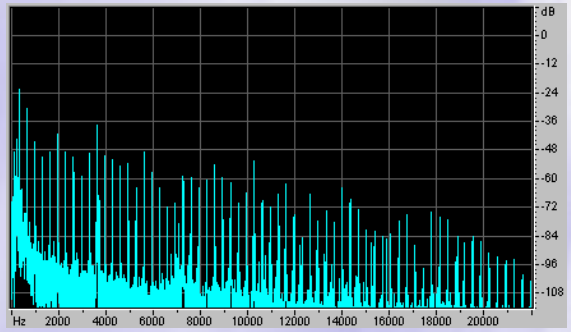
Спектральный анализ

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Примеры звуков и их сонограмм







Нота на гитаре

Форма исходного сигнала

GRAPHICS & MEDIA LAB
VIDEO GROUP

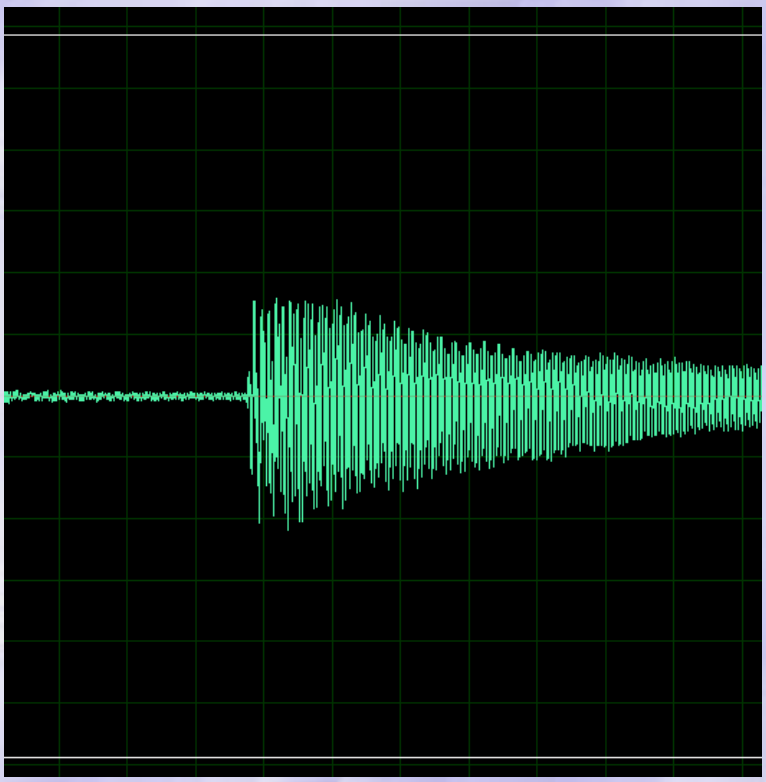
Аудио-сигнал представляют в виде:

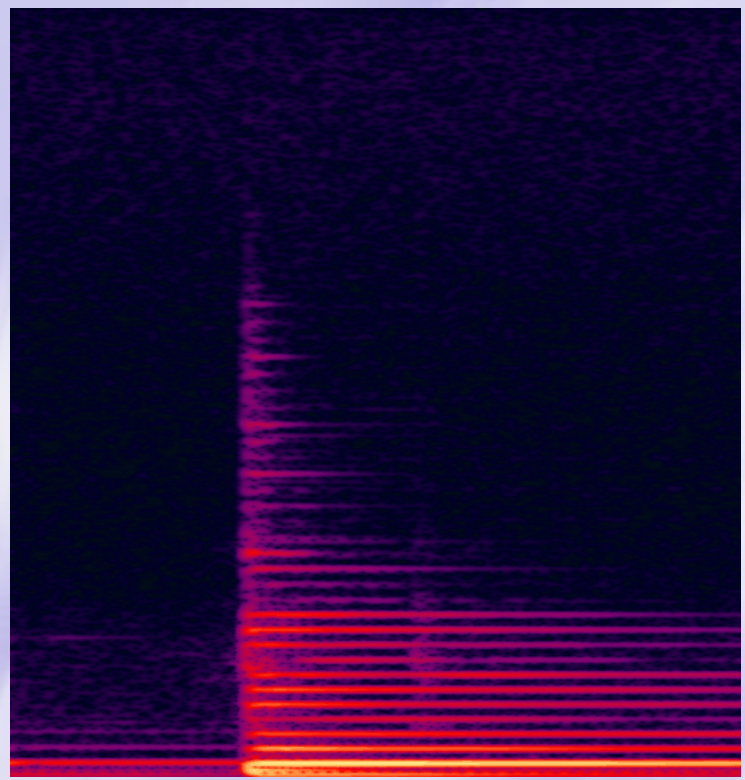
- Набора нот и характеристик инструментов (MIDI)
- Последовательности амплитуд сэмплов (PCM)

При сэмплировании базовые частоты дискретизации от 192 КГц до 6 КГц, точность представления сэмплов – 8, 16, 24, 32 бита.

Качество Audio CD-ROM – 44 КГц, 16 бит.

Пример – гитарная струна

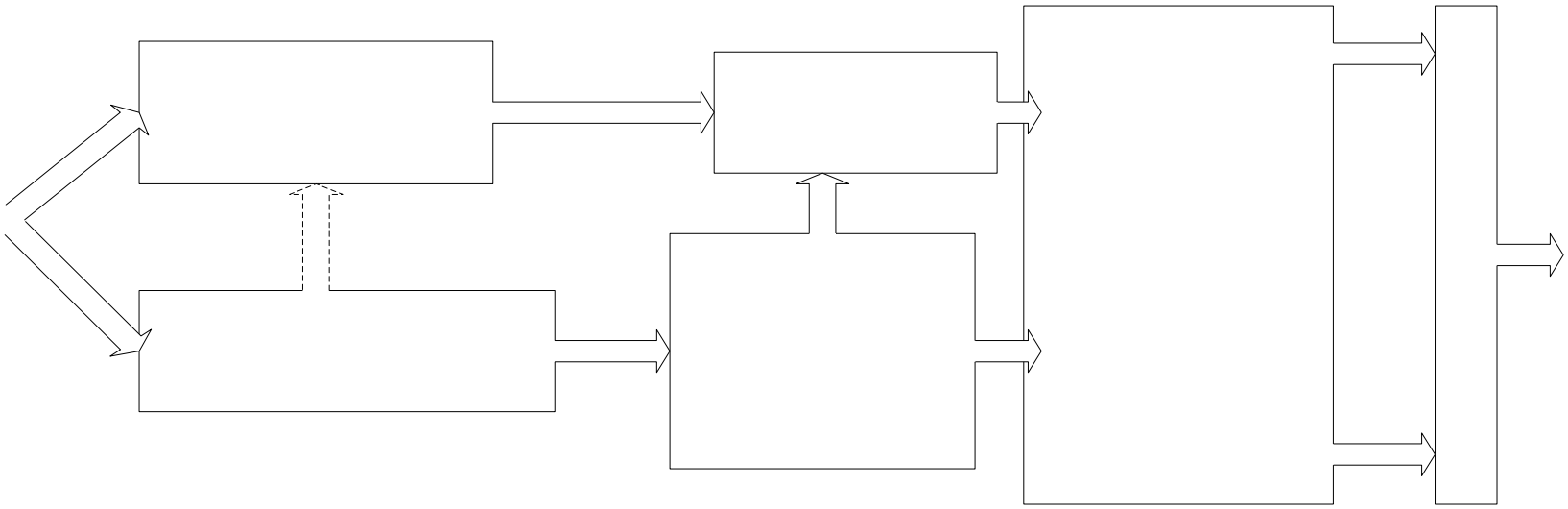




Вид сигнала в виде графика значений амплитуды и в спектральном виде (тот же кусок)

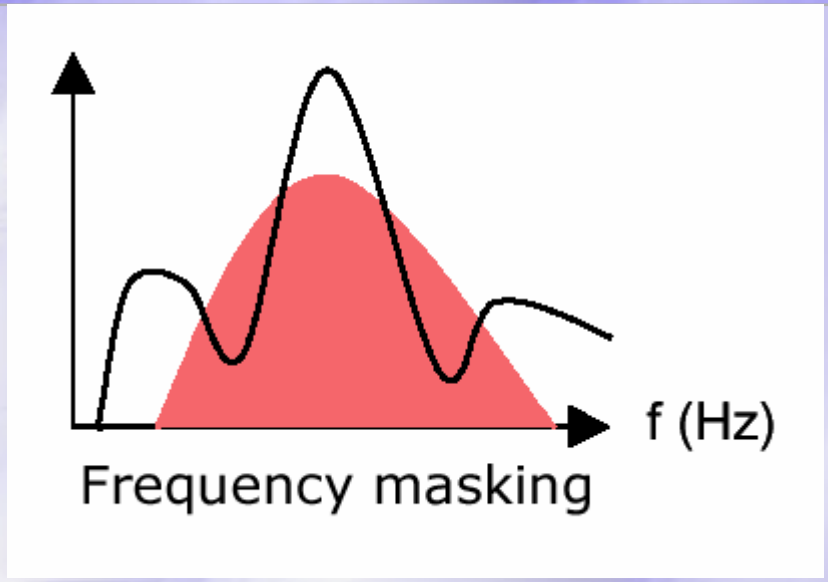
Обобщенная схема аудио-кодека

GRAPHICS & MEDIA LAB
VIDEO GROUP



Подавляющее количество кодеков строится по одной схеме – некая модель (психоаккустика), управляет частотно-временным преобразованием (MDCT, Wavelet), а Rate control – квантованием и энтропийным кодером.

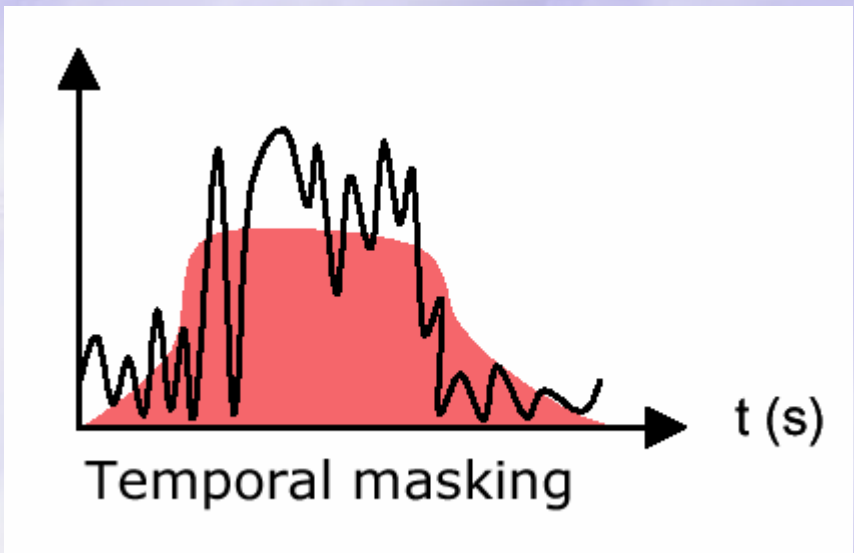
Частотная маскировка



Снижение чувствительности к амплитуде близких по частоте волн в окрестности волны большой амплитуды.

Ухо среднего человека различает порядка 20 частотных полос. При этом в каждой полосе оно чувствительно к тону (у людей со слухом), но реагирует на общую мощность сигнала.

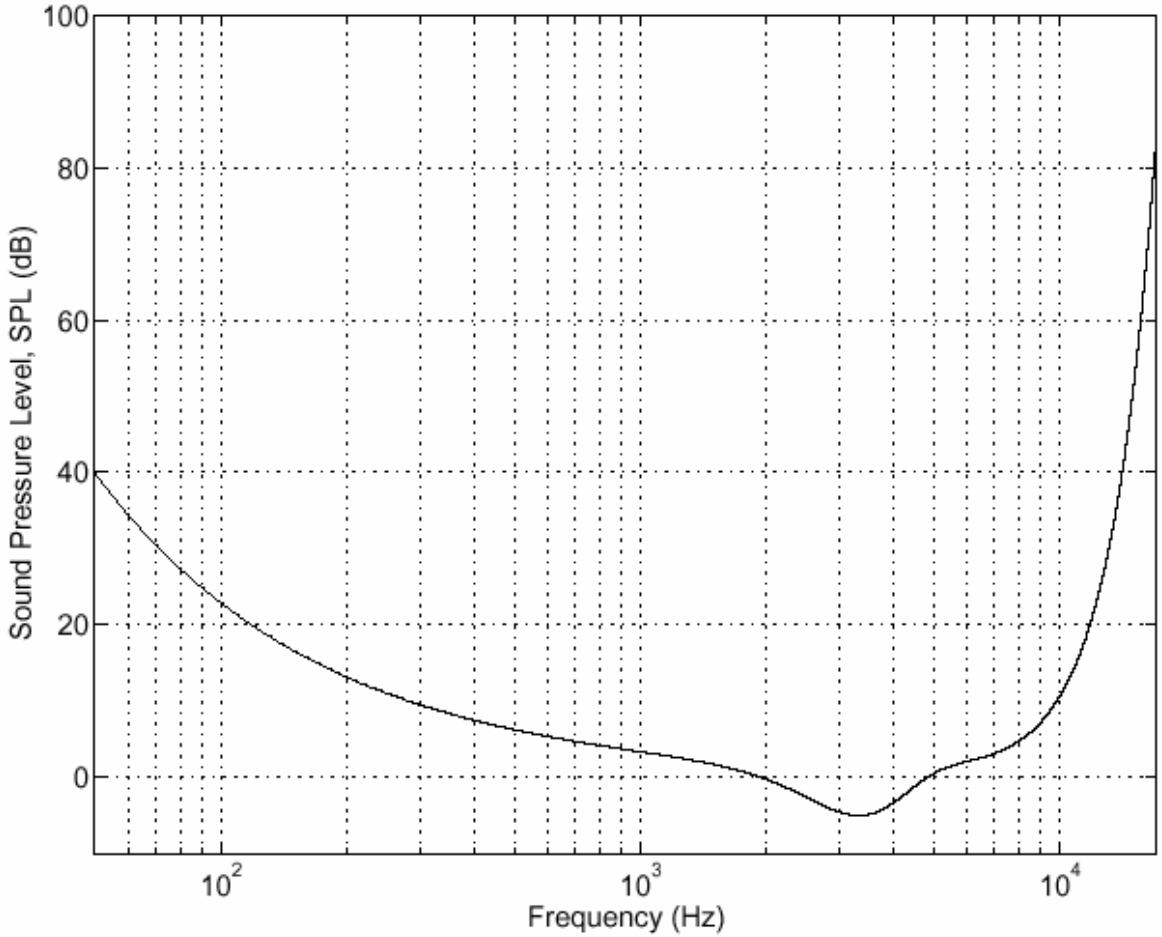
Маскировка по времени



Снижение чувствительности к амплитуде близких по времени волн в после волны большой амплитуды.

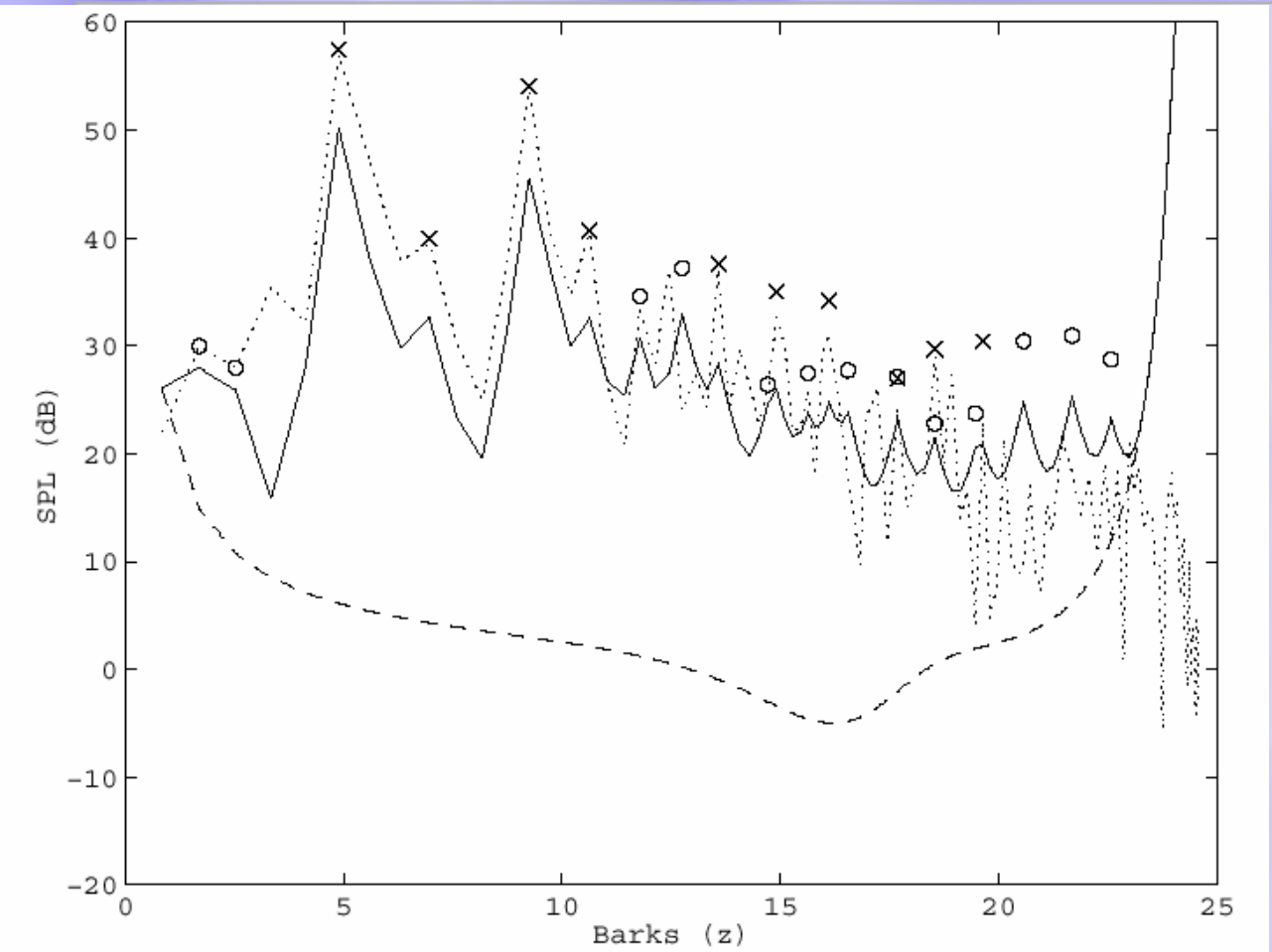
Ухо человека теряет чувствительность («оглушается») волнами большой амплитуды.

Абсолютный порог слышимости



Наивысшая чувствительность уха – на средних слышимых частотах (район 2-3 КГц)

Порог слышимости (психоаккустика)



Психоаккустические пороги определяют по маскировке тоном и шумом, абсолютному порогу слышимости и областям чувствительности

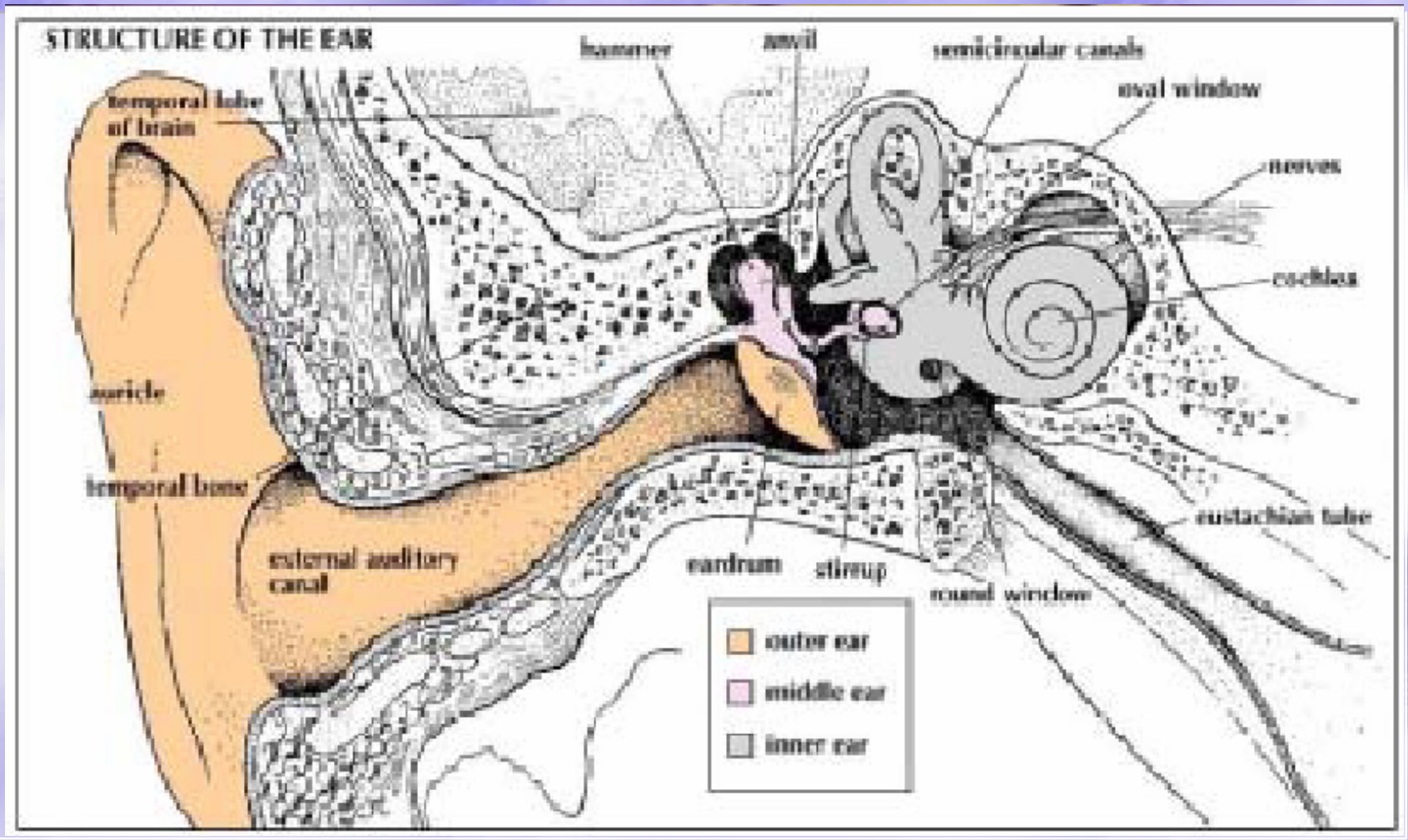
Устройство уха

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Базилярная мембрана
- Различная жесткость мембраны в разных местах → различные резонансные частоты
- К различным участкам мембраны подходят различные группы нервов
- Разложение на частоты

- Описание процессов слухового восприятия в терминах частотно-временной модели

Схема уха


STRUCTURE OF THE EAR
hammer
anvil
semicircular canals
oval window
temporal lobe of brain
nerves
cochlea
auricle
temporal bone
external auditory canal
eustachian tube
eardrum
stirrup
round window
outer ear
middle ear
inner ear

Основные идеи психоаккустики

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Описание свойств слуховой системы человека, на которой основана технология кодирования
- Чувствительность человеческого слуха находится в диапазоне от 2.5 до 5 кГц
- Значимое свойство психоаккустики – эффект маскирования спектральных звуковых элементов
- Неслышимые аудиосигналы несущественны для человеческого восприятия, поэтому могут быть удалены

Психоаккустика

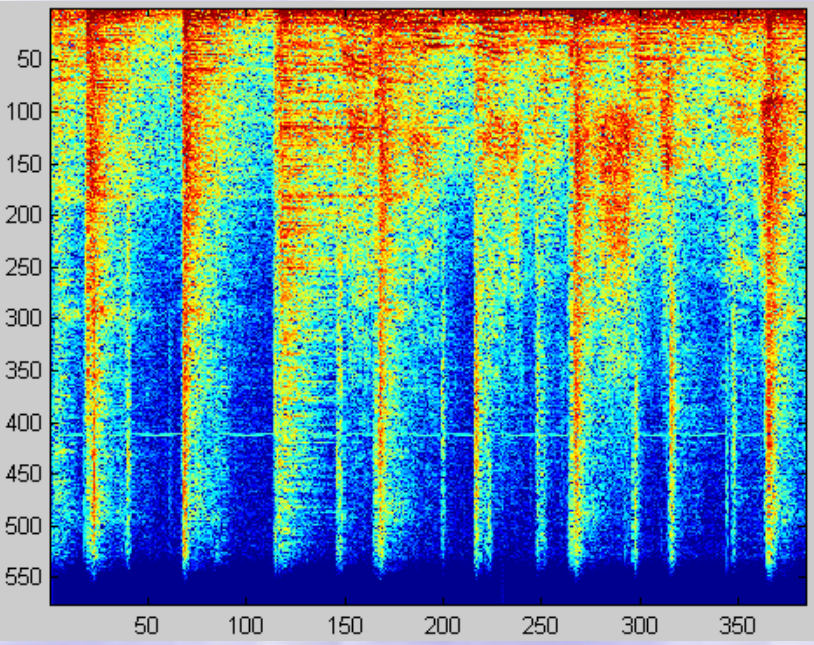
GRAPHICS & MEDIA LAB
VIDEO GROUP

- Построение частотно-временных порогов слышимости шумов квантования в зависимости от исходного аудио-сигнала
- Абсолютные пороги слишимости
- Свойство маскирования

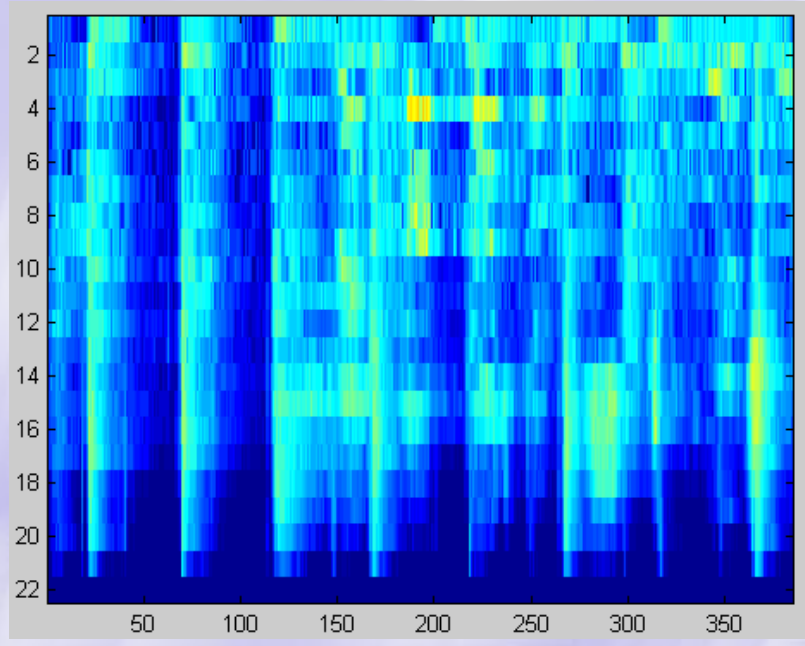
Психоаккустика

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Пример психоакустических порогов



MDCT-спектрограмма



Пороги

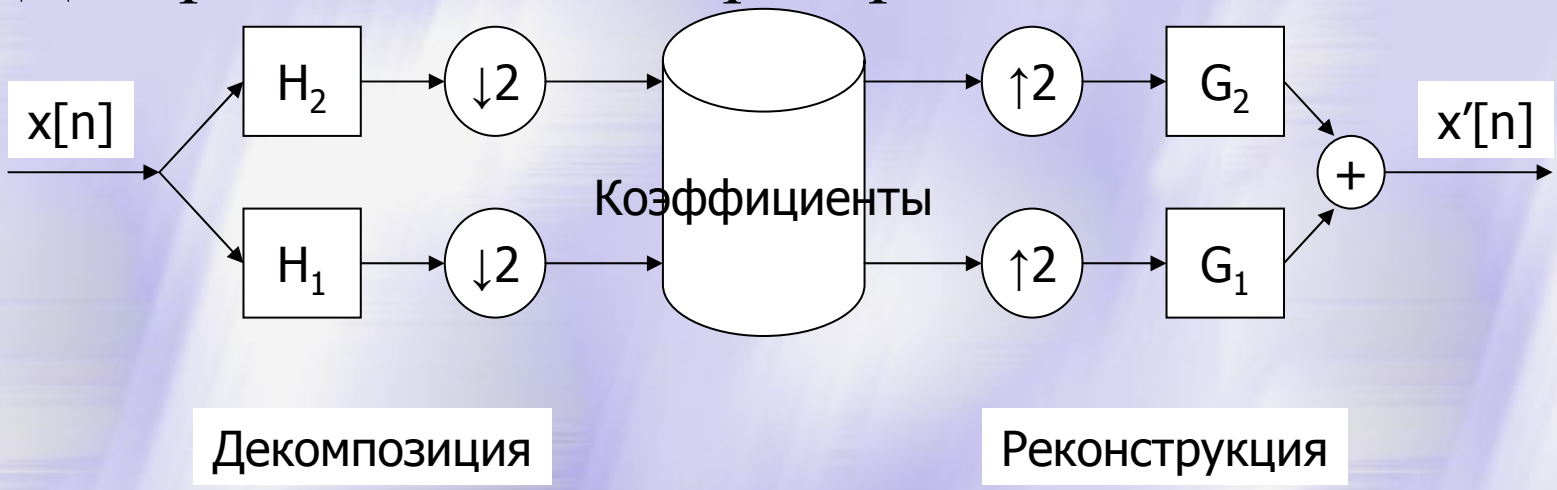
Банки фильтров

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Банк фильтров – преобразование сигнала в несколько сигналов, соответствующих частотным полосам, с возможностью обратного синтеза исходного сигнала.
  - С точным восстановлением?
  - С увеличением количества информации?
  - С гладкими пространственными свойствами?

Вейвлеты как банки фильтров

- Дискретное вейвлет-преобразование

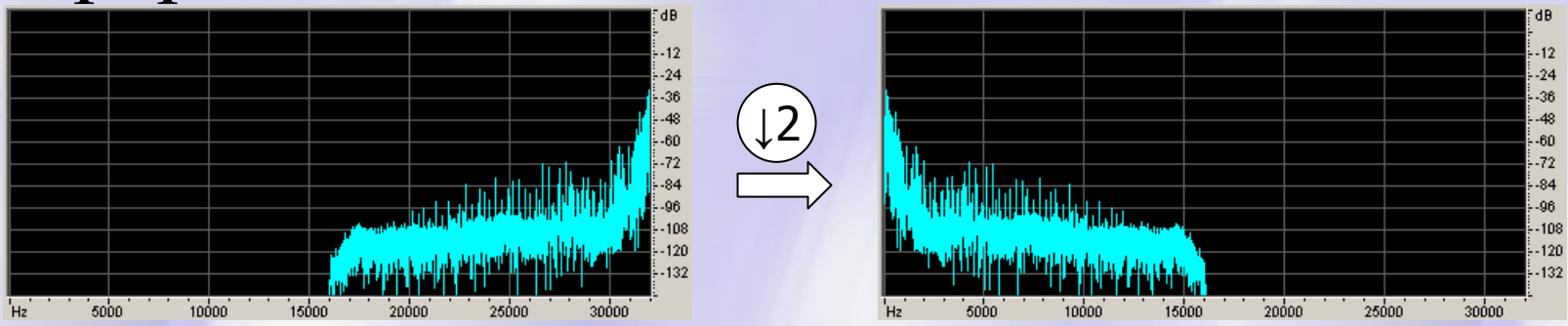


- Свойство точного восстановления (PR):
- Количество информации не изменяется. $x[n] \equiv x'[n]$
- Нужно найти хорошие фильтры, обеспечивающие точное восстановление.

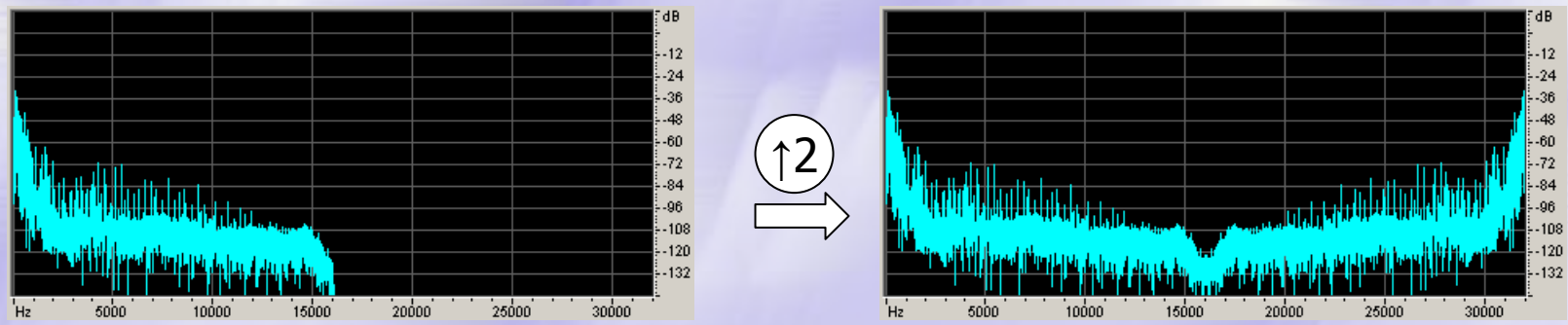
Вейвлеты как банки фильтров

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Прореживание ВЧ-сигнала



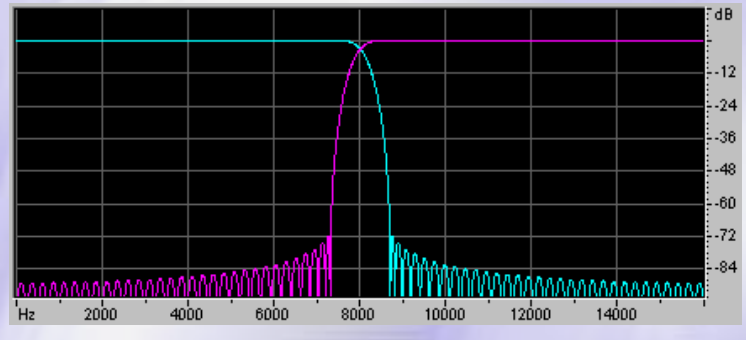
- Интерполяция нулями



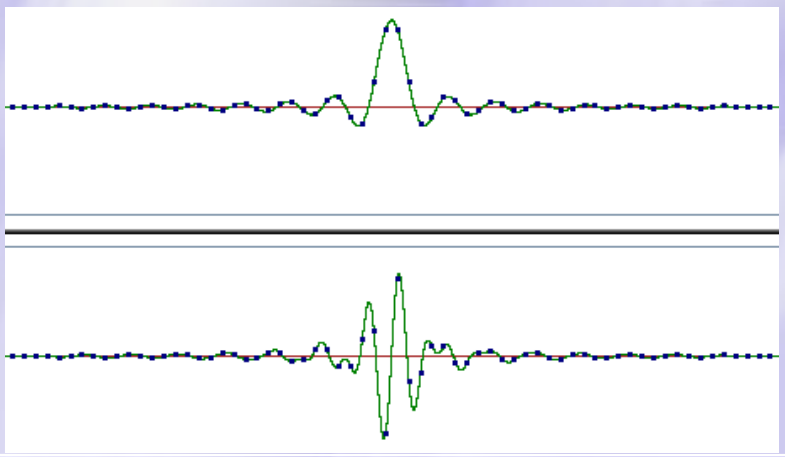
Вейвлеты как банки фильтров

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Квадратурные зеркальные фильтры (QMF)



частотные характеристики

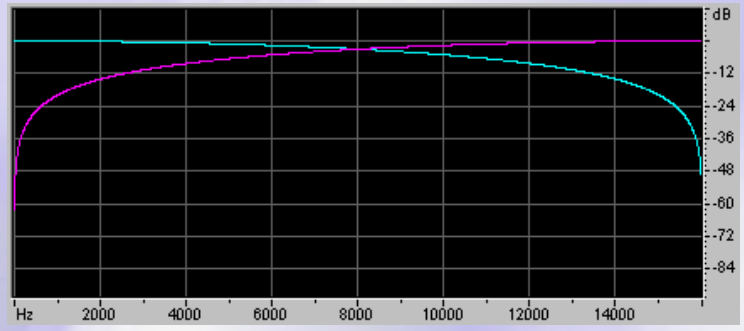


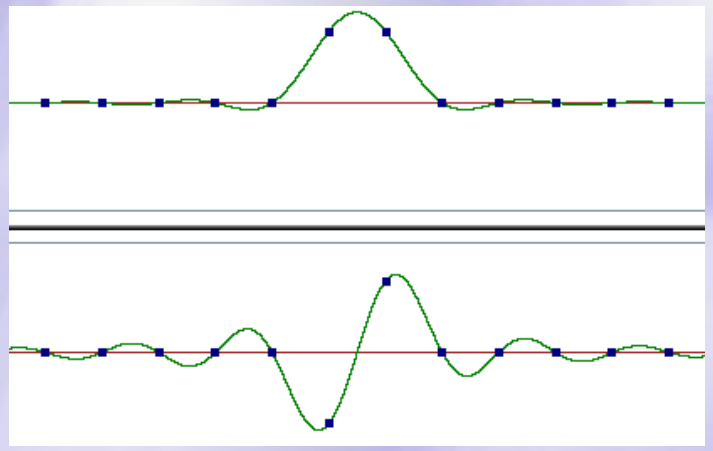
импульсные характеристики

Вейвлеты как банки фильтров

GRAPHICS & MEDIA LAB
VIDEO GROUP

- QMF: базис Хаара

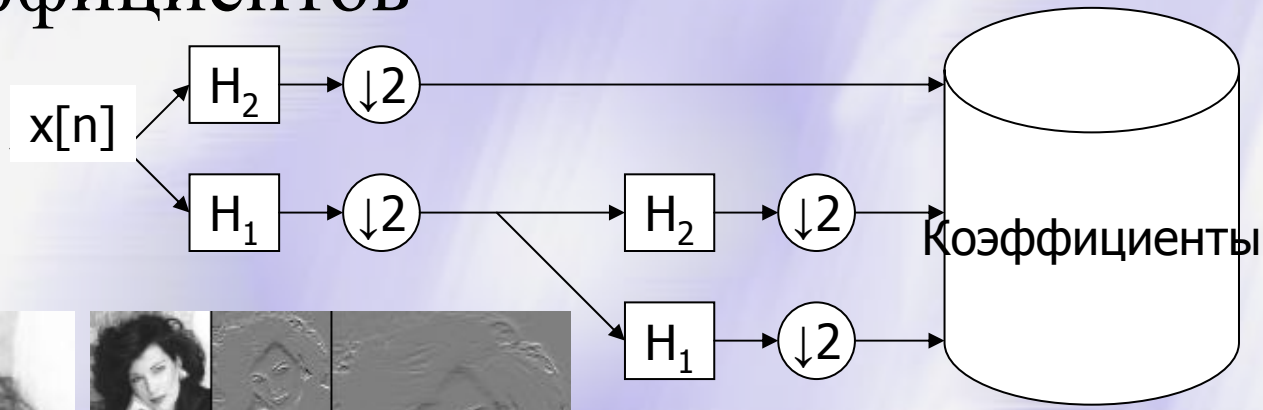




Плохое частотное разделение, но хорошая временная (пространственная) локализация

Пирамидальное представление

- Продолжаем вейвлет-разложение для НЧ-коэффициентов

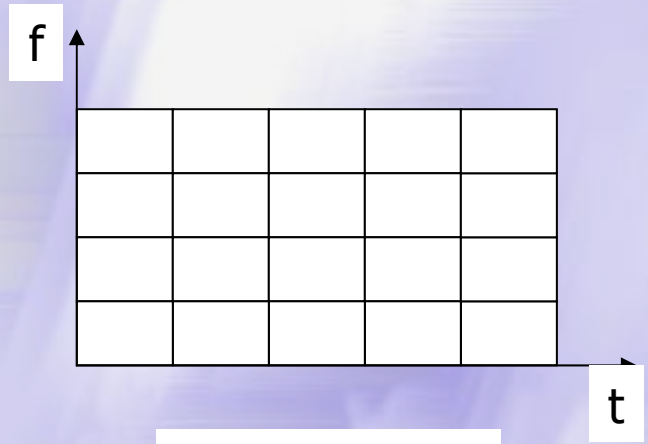


Двумерное вейвлет-преобразование

(на каждом шаге получаем 4 набора коэффициентов)

Банки фильтров

- Как банки фильтров разбивают частотно-временную плоскость?



Оконное ДПФ

Вейвлеты

Банки фильтров: FFT

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Без окон, без перектытия
  - Плохое разделение по частотам
  - Временной алиасинг
  - Нет избыточности
- С окнами, с перекрытием
  - Хорошее разделение по частотам
  - Нет временного алиасинга (при двукратном применении окон)
  - Избыточность

Банки фильтров: MDCT

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Хорошее разделение по частотам
- С перекрытием и уничтожением временного алиасинга
- Без избыточности!

Каждое окно длины 2N захватывает N новых отсчетов и выдает N коэффициентов.

Банки фильтров: MDCT

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Входные блоки: 2N точек, из них только N новых
- Выходные коэффициенты:

N действительных коэффициентов на блок

- Весовые окна:

$$h^2[n] + h^2[N-1-n] = 2, \quad 0 \le n < N$$

$$h^2[N+n] + h^2[2N-1-n] = 2, \quad 0 \le n < N$$

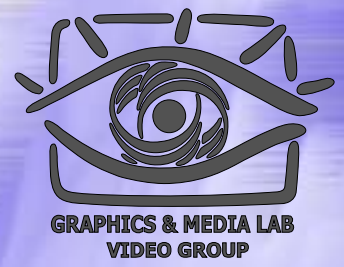
Избыточность аудио

GRAPHICS & MEDIA LAB
VIDEO GROUP

В аудио-сигнале избыточны:

- Амплитуды тонов, вблизи всплесков тонов (маскировка по частоте)
- Амплитуды сигнала после всплесков сигнала (маскировка по времени)
- Низкие и высокие частоты могут быть представлены менее точно
- Разные каналы в стерео и 5.1 могут быть весьма похожи

Использование стандарта аудиокодирования MPEG

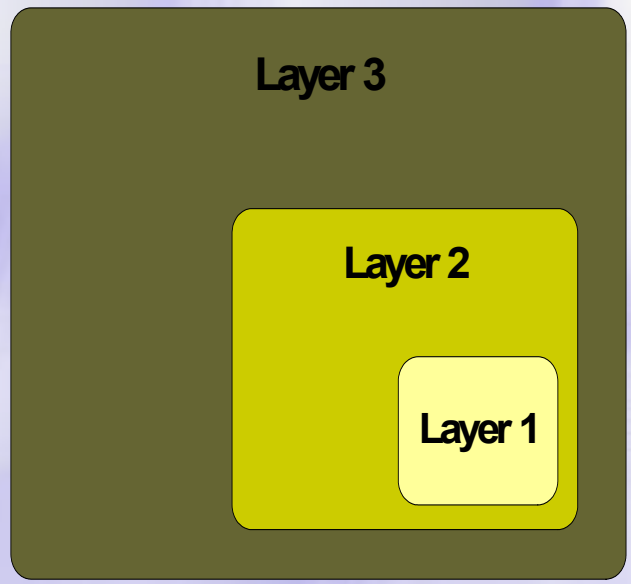


- Цифровое аудиовещание (EUREKA DAB, WorldSpace, ARIB, DRM)
- Передача данных в сетях ISDN
- Архивное хранение эфирных материалов
- Звуковая дорожка в цифровом телевидении (DVB, Video CD, ARIB)
- Потоковые медиаданные в интернете (Microsoft Netshow, Apple Quicktime)
- Портативные плееры (mpman, mplayer3, Rio, Lyra,YEPP, iRiver и др.)
- Хранение и перенос музыкальных файлов

MPEG-1 Audio

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Первая фаза разработки группы MPEG. Началась в 1988 и закончилась в конце 1992 выработкой стандарта ISO/IEC IS 11172
- MPEG-1 состоит из трех уровней повышающейся сложности кодирования



MPEG-1 Audio

GRAPHICS & MEDIA LAB
VIDEO GROUP

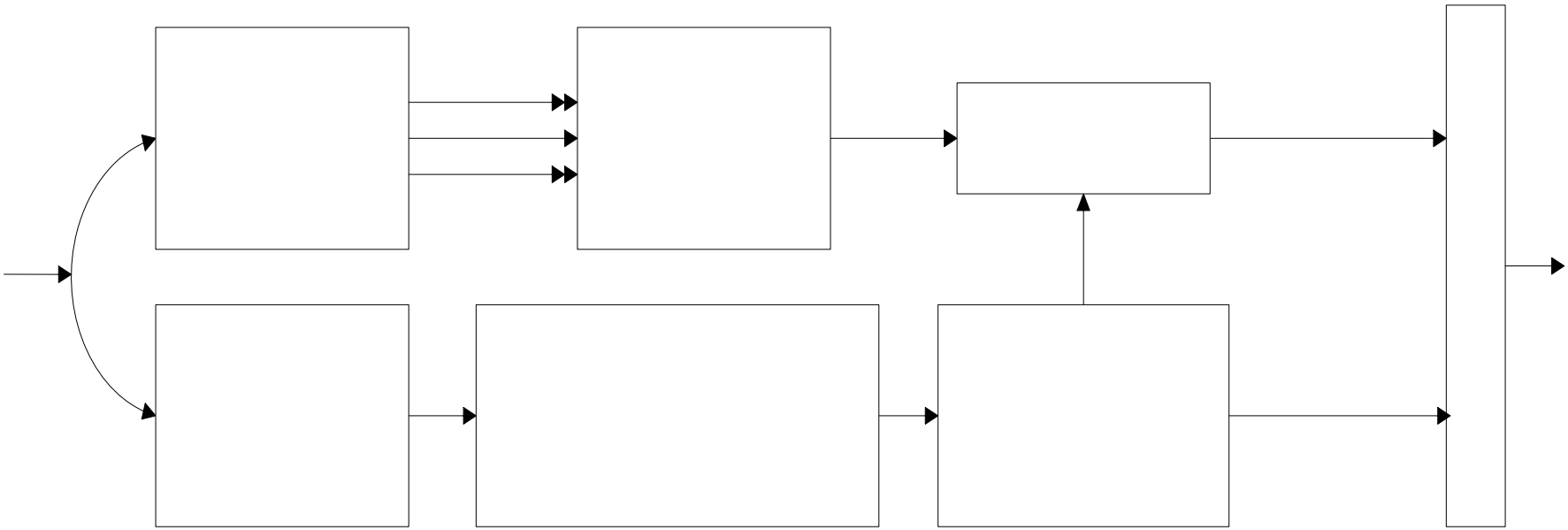
ISO/IEC 11172-3 (MPEG-1): 1992

- Вход: 16-bit PCM, 32, 44.1 и 48 кГц
- Выход:
  - mono, stereo, dual independent mono и joint stereo
  - rate: 32-196 Кб/с (mono), 64-384 Кб/с (stereo)
- MPEG-1 layer III: MP3

MPEG-1

Диаграммы кодирования

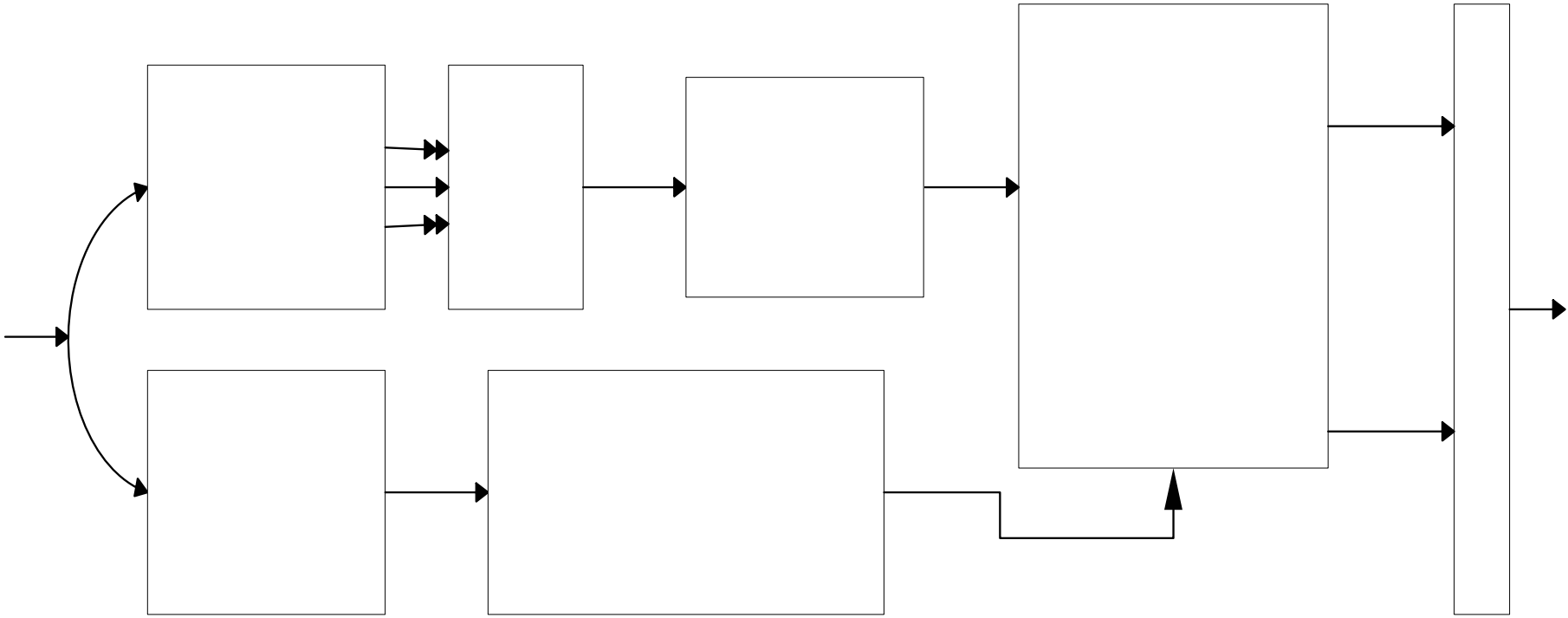
- MPEG1 – I/II



MPEG-1

Диаграммы кодирования

- MPEG1 – III (MP3)



MPEG-2 Audio

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Были внедрены новые понятия в MPEG кодировании видео, такие как чересстрочные видеосигналы.
- Основная область применения MPEG-2 – это цифровое телевидение
- Законченный в 1994 году стандарт MPEG-2, состоит из двух расширений MPEG-1, не предложивших новых алгоритмов кодирования

MPEG-2 Audio

GRAPHICS & MEDIA LAB
VIDEO GROUP

ISO/IEC 13818-3 BC/LSF:

- Не исключает использование прежних версий
- Поддержка низких частот
- Кодирование стереосигналов. Известная из звуковых дорожек к фильмам конфигурация “5.1 - аудио”
- Поддержка mono, stereo 16, 22.05, 24, 32, 44.1 и 48 кГц
- Битрейт: 32-640 Кб/с

MPEG-2 AAC

GRAPHICS & MEDIA LAB
VIDEO GROUP

В 1994 году проверочный тест показал, что предложенные новые алгоритмы кодирования (без обратной совместимости с MPEG-1) значительно повысят эффективность кодирования.

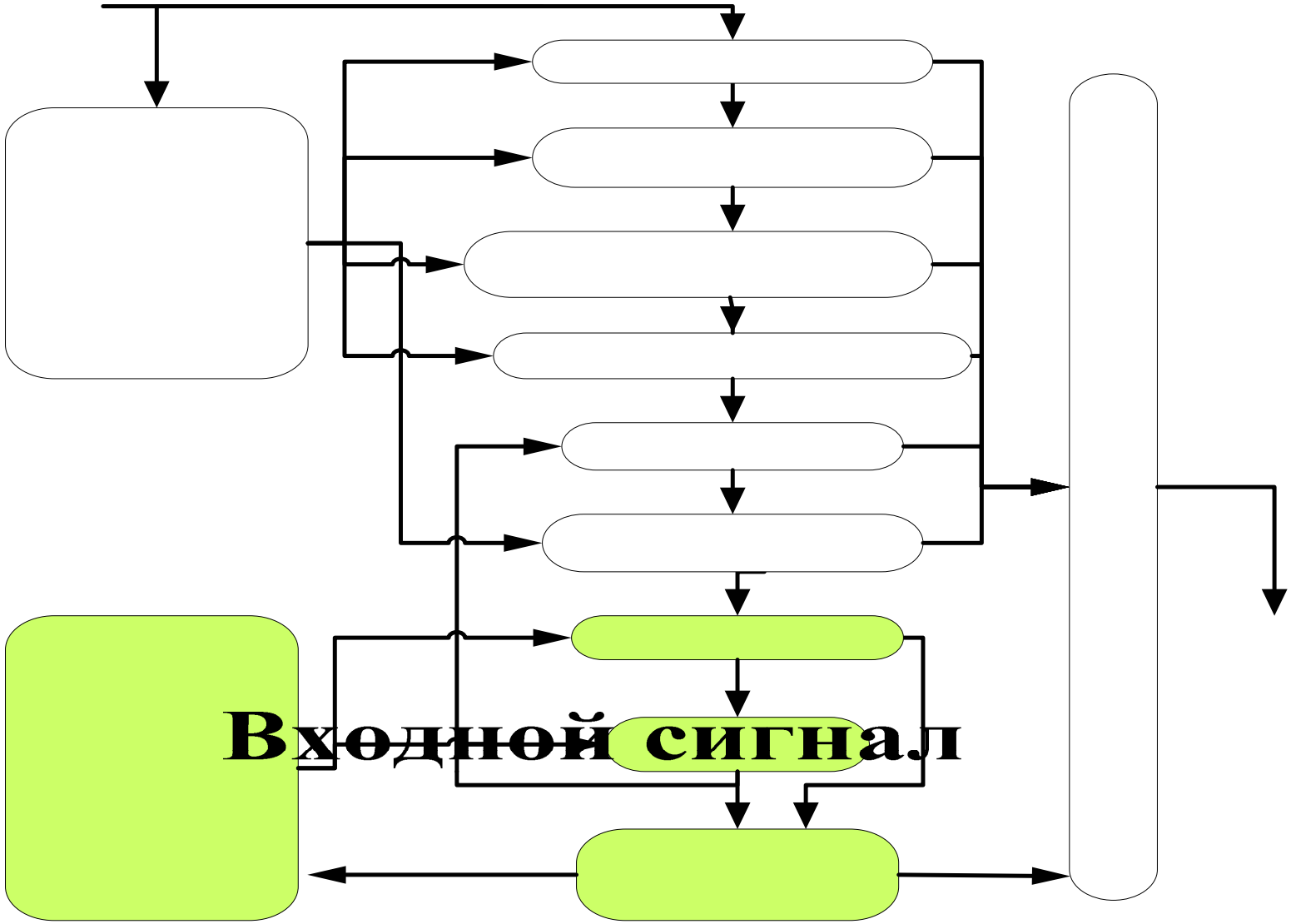
Так появился MPEG-2 Advanced Audio Coding (AAC)

MPEG-2
Advanced Audio Coding

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Основной принцип ACC кодирования аналогичен Layer-3.
- ACC имеет ряд улучшений в некоторых деталях. Использует новые средства для улучшения качества кодирования при низких битрейтах.

Схема кодирования AAC



MPEG-2 AAC

GRAPHICS & MEDIA LAB
VIDEO GROUP

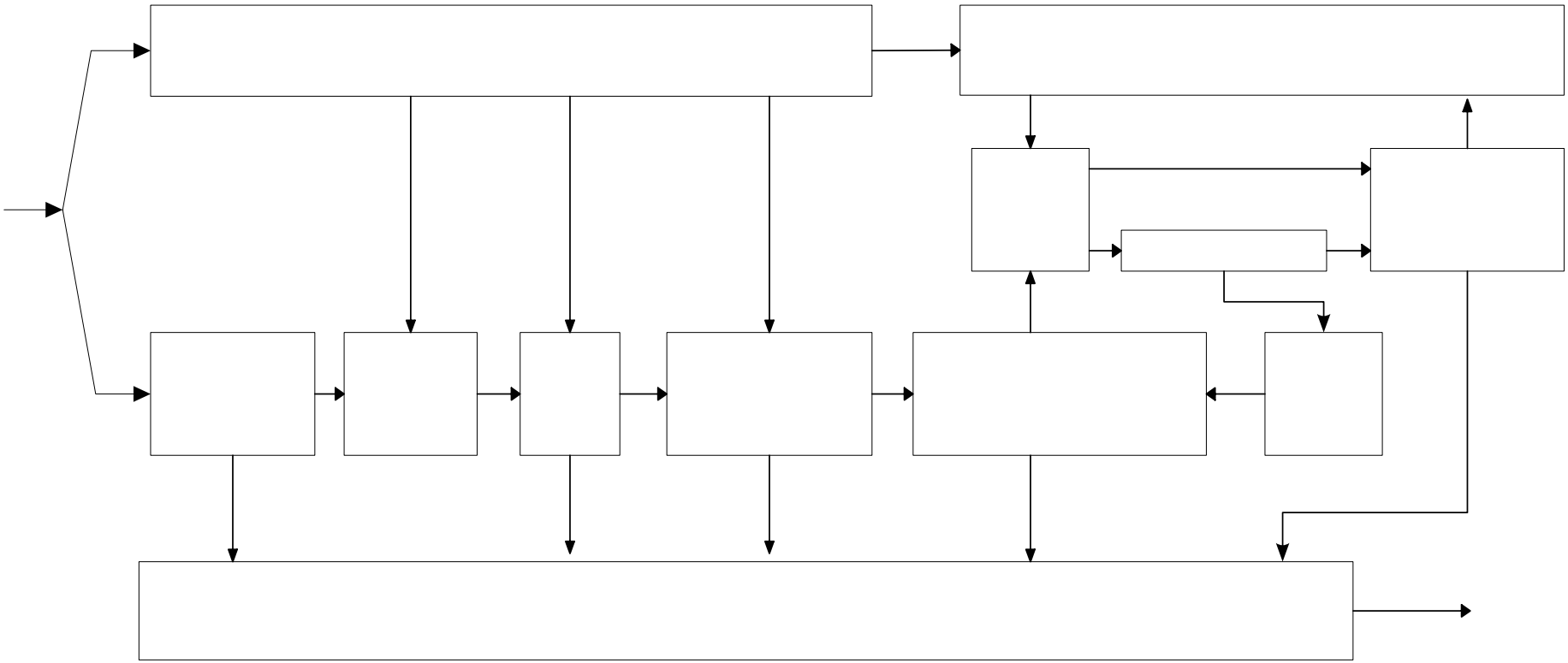
- ISO/IEC 13818-7 NBC/AAC:
  - NBC/AAC: Несовместим с прежними версиями / расширенное кодирование звука
  - 5 каналов: левый, правый, центральный, окружающий левый, окружающий правый
  - Поддержка 32, 44.1 и 48 кГц
  - Частота 8-64 Кб/с на канал

MPEG-2

GRAPHICS & MEDIA LAB
VIDEO GROUP

- MPEG-2 NBC/AAC

диаграмма кодирования



MPEG-2 AAC
Audio Transport Formats

GRAPHICS & MEDIA LAB
VIDEO GROUP

- ADIF (Audio Data Interchange Format)

  Все данные контроля декодера (частота семплирования, режим и т.д.) помещаются в один заголовок, идущий перед аудиопотоком. Не позволяет кодировать начиная с определенной точки, как в стандарте MPEG-1
- ADTS (Audio Data Transport Stream)

  Пакует AAC-данные во фреймы с заголовками очень похожими на заголовки MPEG1/2. Позволяет кодировать начиная с середины потока.

Почему MP3?

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Открытый стандарт
- В течение многих лет существуют аппаратные и программные кодировщики и декодировщики
- Поддерживается многими технологиями
- Короче, MP3 – нужная технология, ставшая доступной в нужное время

MPEG-1/2 Layer-3

GRAPHICS & MEDIA LAB
VIDEO GROUP

В заголовке файла указывается:

- Слово синхронизации
- Битрейт
- Частота семплирования
- Layer
- Режим кодирования
- SCMS (Serial Copy Management Scheme)

MP3
Гибкость применения

GRAPHICS & MEDIA LAB
VIDEO GROUP

Рабочий режим

- Один канал
- Два независимых канала
- Stereo
- Joint stereo

MP3
Гибкость применения

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Частота дискретизации
  - MPEG-1: 32, 44.1 и 48 кГц
  - MPEG-2: 16, 22.5 и 24 кГц
  - MPEG-2.5 (расширение MP3): 8, 11.05 и 12 кГц
- Скорость передачи битов
  - Поддерживается переменная и постоянная скорость передачи битов

MP3: Введение

GRAPHICS & MEDIA LAB
VIDEO GROUP

MPEG-1 layer-III (более широко известный как MP3) – был стандартизован в 1991 в рамках кодирования видео Moving Pictures Expert Group ISO (образована в 1988).

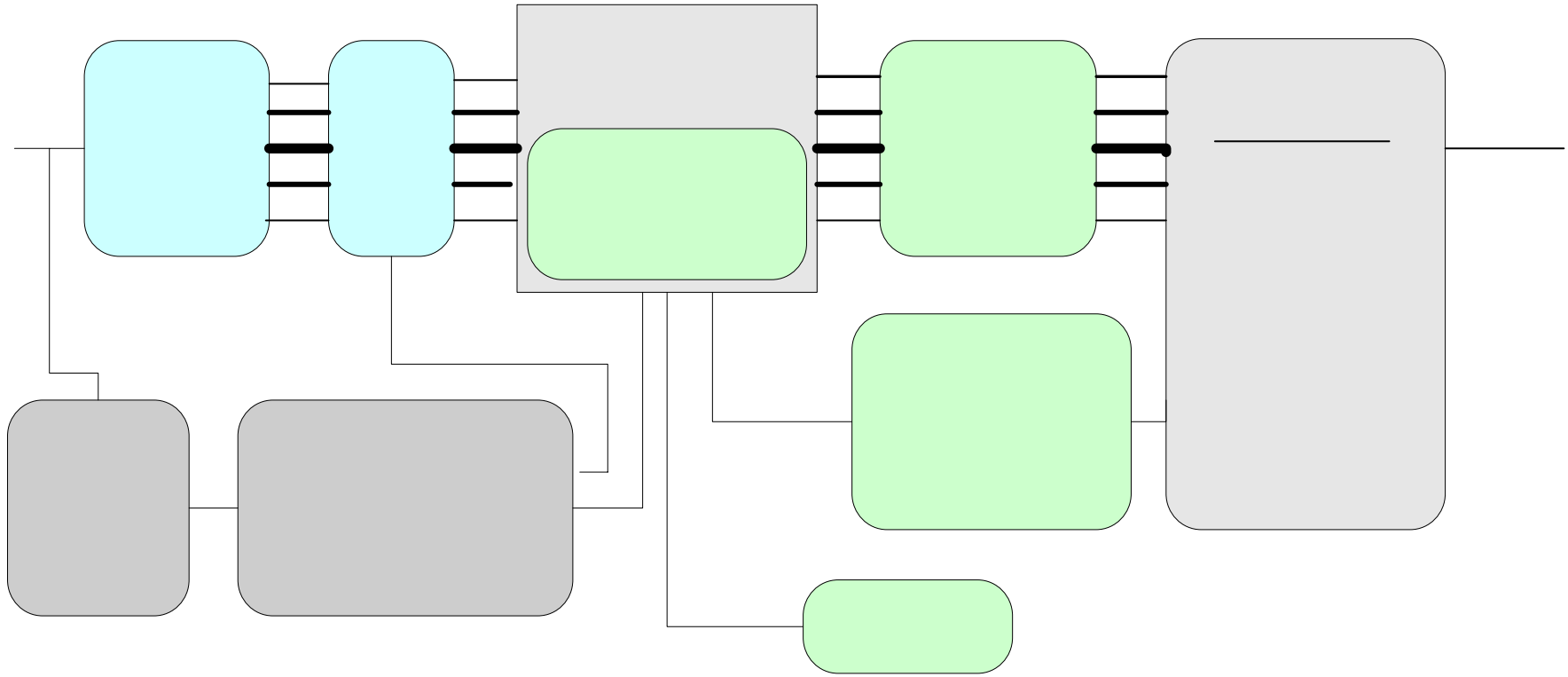
Стандарты MPEG ориентированы в т.ч. на аппаратную реализацию алгоритмов (используется сейчас в DVD и спутниковом телевидении). Включает 3 уровня сложности алгоритма I, II, III. Layer-III – самый сложный.

MP3: Введение (2)

В 1993 году, с разработкой стандарта MPEG-2, MP3 был расширен:

- Добавлена поддержка до 6 каналов (звук 5.1).
- Добавлена поддержка низких частот сэмплирования входных сигналов, что позволило повысить качество на низких битрейтах

MP3: Общая схема MP3



MP3: Схема работы

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Модифицированное косинусное преобразование (MDCT) используется для разложения сигнала.
- БПФ (FFT) – для расчета психоаккустики
- кодирование по Хаффману — для сжатия без потерь.

MP3: MDCT

Ключевым преобразованием в MP3 является MDCT (суть та же, что в DCT в JPEG & MPEG-4, но преобразование берется с пересекающимися окнами). Прямое и обратное преобразование:

$$y(k) = 2 \cdot \sum_{n=0}^{N-1} x(n) \cos\left(\frac{\pi}{N}(n+n_0)(2k+1)\right), \quad \text{for } 0 \le k < \frac{N}{2}$$

$$x(n) = \frac{2}{N} \cdot \sum_{k=0}^{\frac{N}{2}-1} y(k) \cos\left(\frac{\pi}{N}(n+n_0)(2k+1)\right), \quad \text{for } 0 \le n < N$$

MP3: Квантование (1)

Общий смысл квантования – понижение точности представления данных, причем в аудио это делается на разную величину для разных амплитуд данных:

$$ix(i) = nint\left(\left(\frac{|xr(i)|}{\sqrt[4]{2^{qquant + quantanf}}}\right)^{0.75} - 0.0946\right)$$

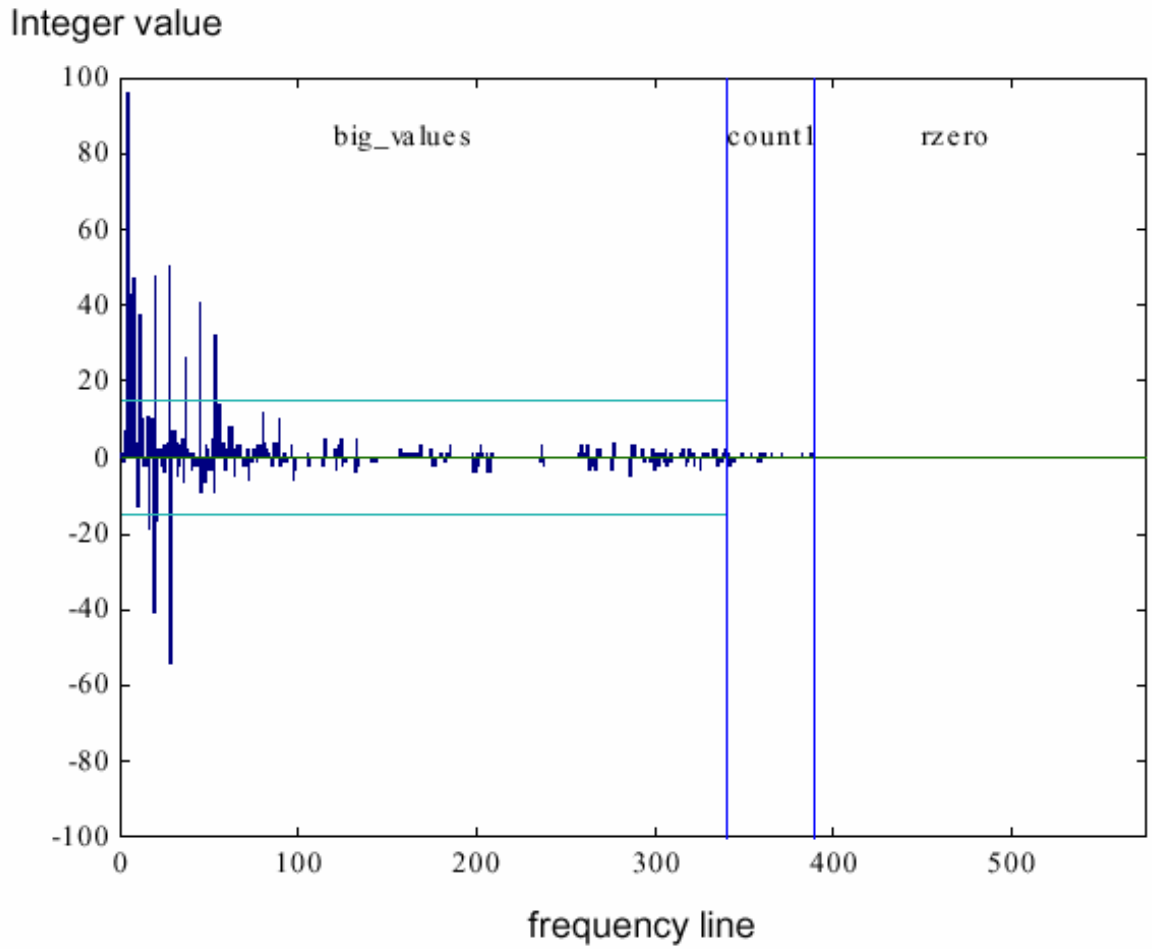
где, ***xr(i)*** – исходные данные, ***qquant*** и ***quantanf*** – значения кванта для всего преобразования и конкретного участка, ***nint()*** – округление к ближайшему целому.

MP3: Квантование (2)

Управляя квантованием – можно:

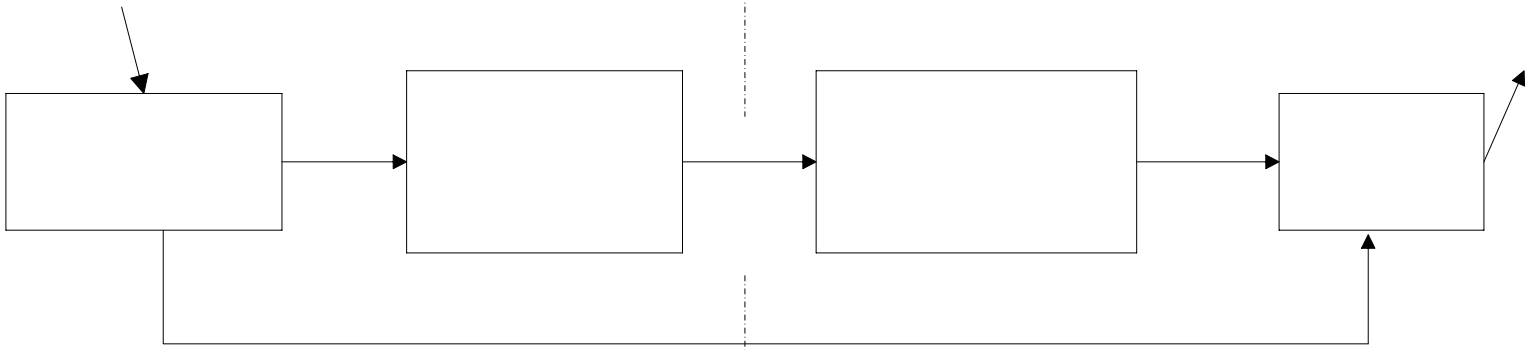
- Задавать точность представления участков спектра (использовать психоакустику для увеличения сжатия),
- Задавать качество участков мелодии (задавать разные стратегии управления размером – CBR, VBR и т.д.)
- Управлять общим размером мелодии (задавать битрейт)

MP3: Распределение амплитуд частот



В реальных мелодиях большая амплитуда у низких частот и малая у высоких

MP3: Квантование и энтропийное сжатие



Данные после MDCT преобразования подвергаются огрублению (от которого зависит битрейт и собственно качество), и далее без потерь сжимаются по Хаффману с фиксированными таблицами.

MP3: Удаление избыточности каналов

GRAPHICS & MEDIA LAB
VIDEO GROUP

Стерео-сигнал может кодироваться 3 способами:

- Независимое сжатие каналов
- Использование MS stereo
- Использование Intensity Processing

MP3: MS stereo

$$L_i = \frac{M_i + S_i}{\sqrt{2}} \quad \text{and} \quad R_i = \frac{M_i - S_i}{\sqrt{2}}$$

M_i – сумма значений в 2 каналах

S_i – разность значений в 2 каналах

Это наиболее простой способ уменьшения избыточности между двумя каналами.

MP3: Intensity stereo

GRAPHICS & MEDIA LAB
VIDEO GROUP

$$is\_ratio = \tan\left(is\_pos_{sb} \cdot \frac{\pi}{12} \right)$$

$$L_i = L_i \cdot \frac{is\_ratio}{1 + is\_ratio}$$

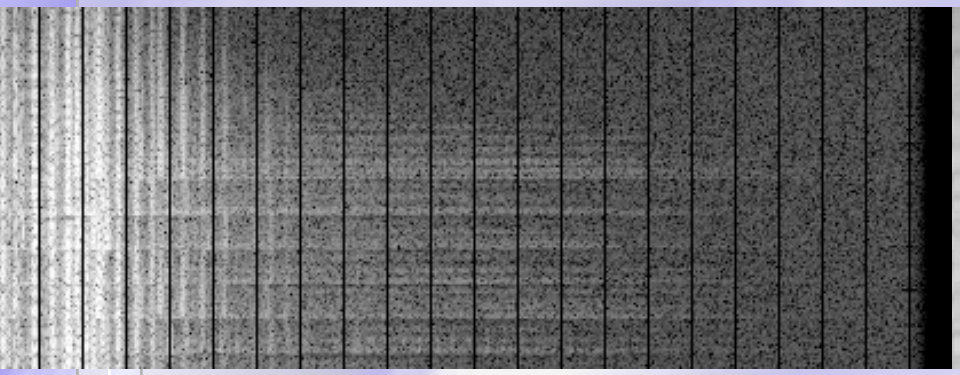
$$R_i = L_i \cdot \frac{1}{1 + is\_ratio}$$

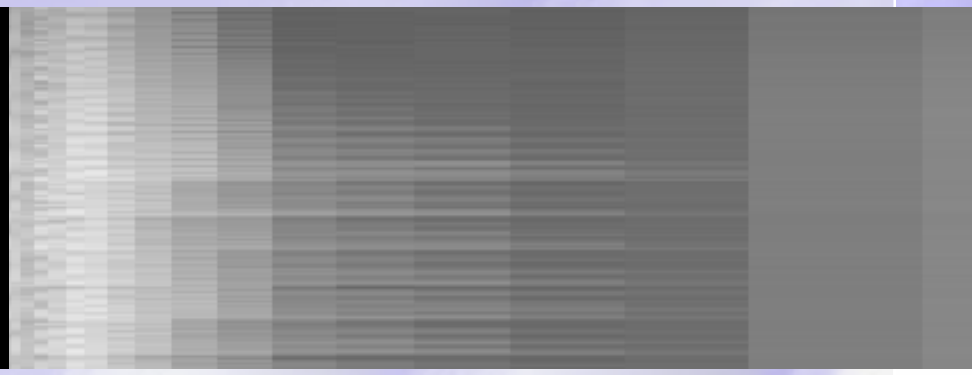
Метод работает, когда части спектра стерео-сигнала пропорциональны (как правило на высоких частотах)

Более сложный метод, использующий разложение амплитуду и стерео часть, и сохраняющий данные is_pos_{sb} в данных коэффициентов квантования.

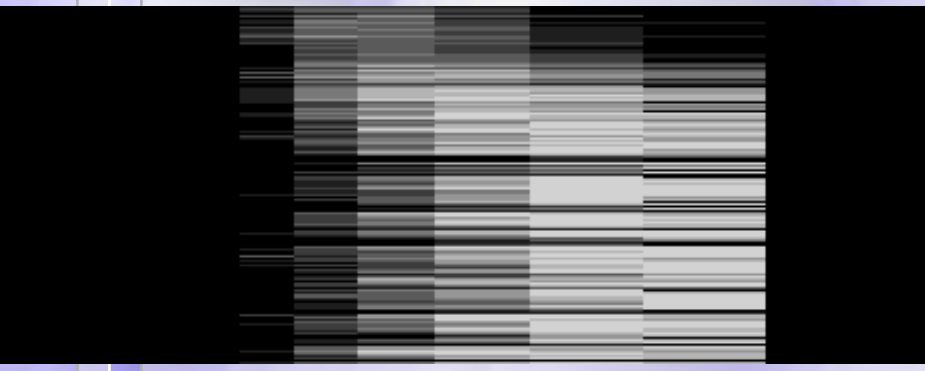
Приведены формулы восстановления сигнала.

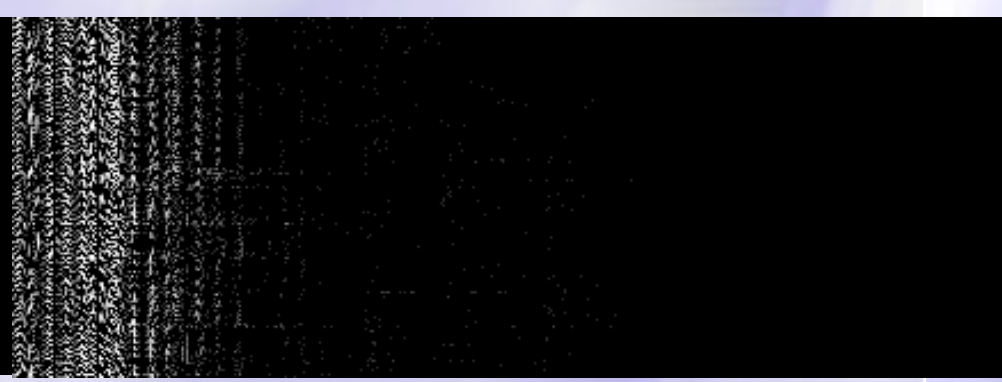
MP3: Визуализация значений коэффициентов





Исходные данные и пороги чувствительности





кванты и отквантованные значения

Схема преобразований MP-3

GRAPHICS & MEDIA LAB
VIDEO GROUP

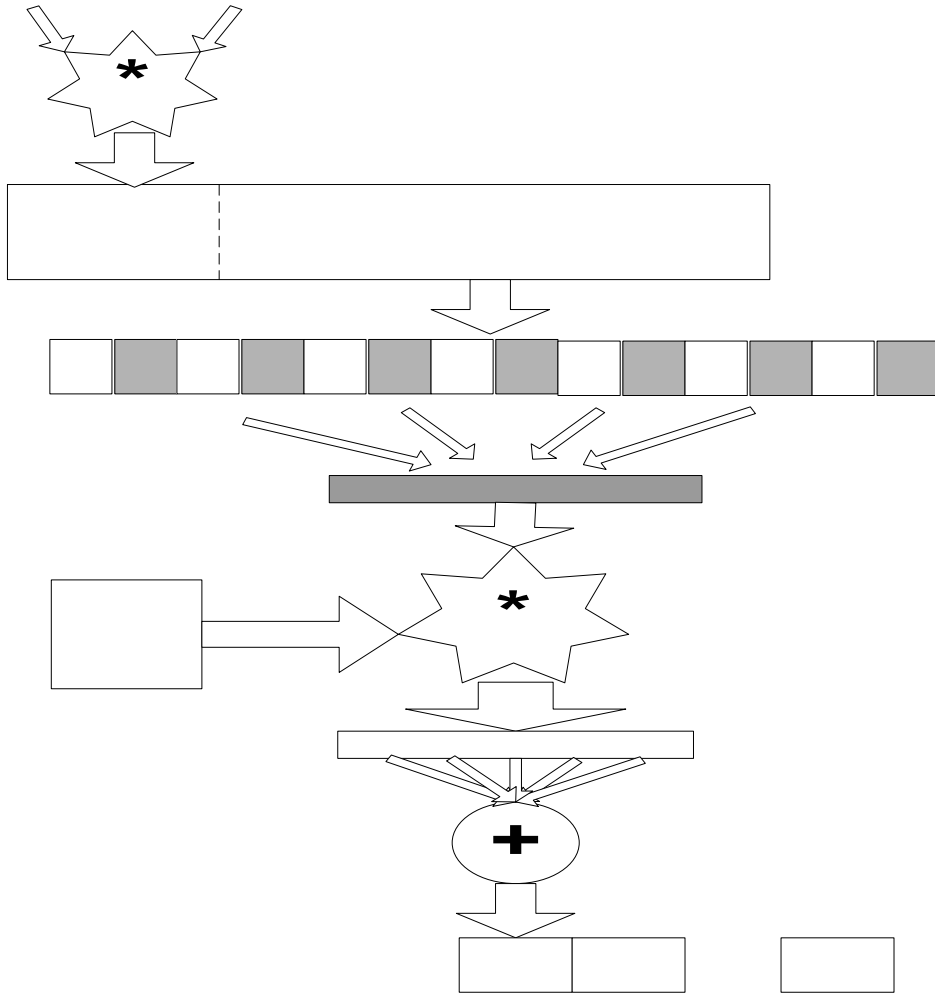
$$Y_i = \sum_{k=0}^{31} N_{ik} S_k; \qquad 0 \le i < 64$$

```
for i=1023 downto 64 do V[i]=V[i-64]
for I=0 downto 63 do   V[i]=Y[i]

for i=0 to 7 do
  for j=0 to 31 do
    U[i*64+j]=V[i*128+j]
    U[i*64+32+j]=V[i*128+96+j]
```

$$W_i = U_i D_i; \qquad 0 \le i < 512$$

$$S_j = \sum_{i=0}^{15} W_{j+32i}; \qquad 0 \le j < 32$$



Применение

GRAPHICS & MEDIA LAB
VIDEO GROUP

- MPEG-1 layer I: 384 kb/s, цифровые кассеты (DCC)
- MPEG-1 layer II: 224 kb/s, цифровое спутниковое вещание (DBS)
- MPEG-1 layer II: 256 kb/s, трансляция цифровой звукозаписи, Eureka 147 digital
- MPEG-1layer III: MP3
- MPEG-2 BC/LSF: кино
- MPEG-2 NBC/AAC: Internet, LiquidAudio, DRM, Xradio.

MPEG-3

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Планировалось определить стандарты кодирования видео высокой четкости (HDTV) и назвать их MPEG-3. Но до этого было решено, что возможности MPEG-2 вполне подходят для HDTV. Таким образом разработки MPEG-3 были включены в MPEG-2. В результате от MPEG-3 отказались в пользу MPEG-4

- Не путать MPEG-1/2 Layer-3 (MP3) с MPEG-3!

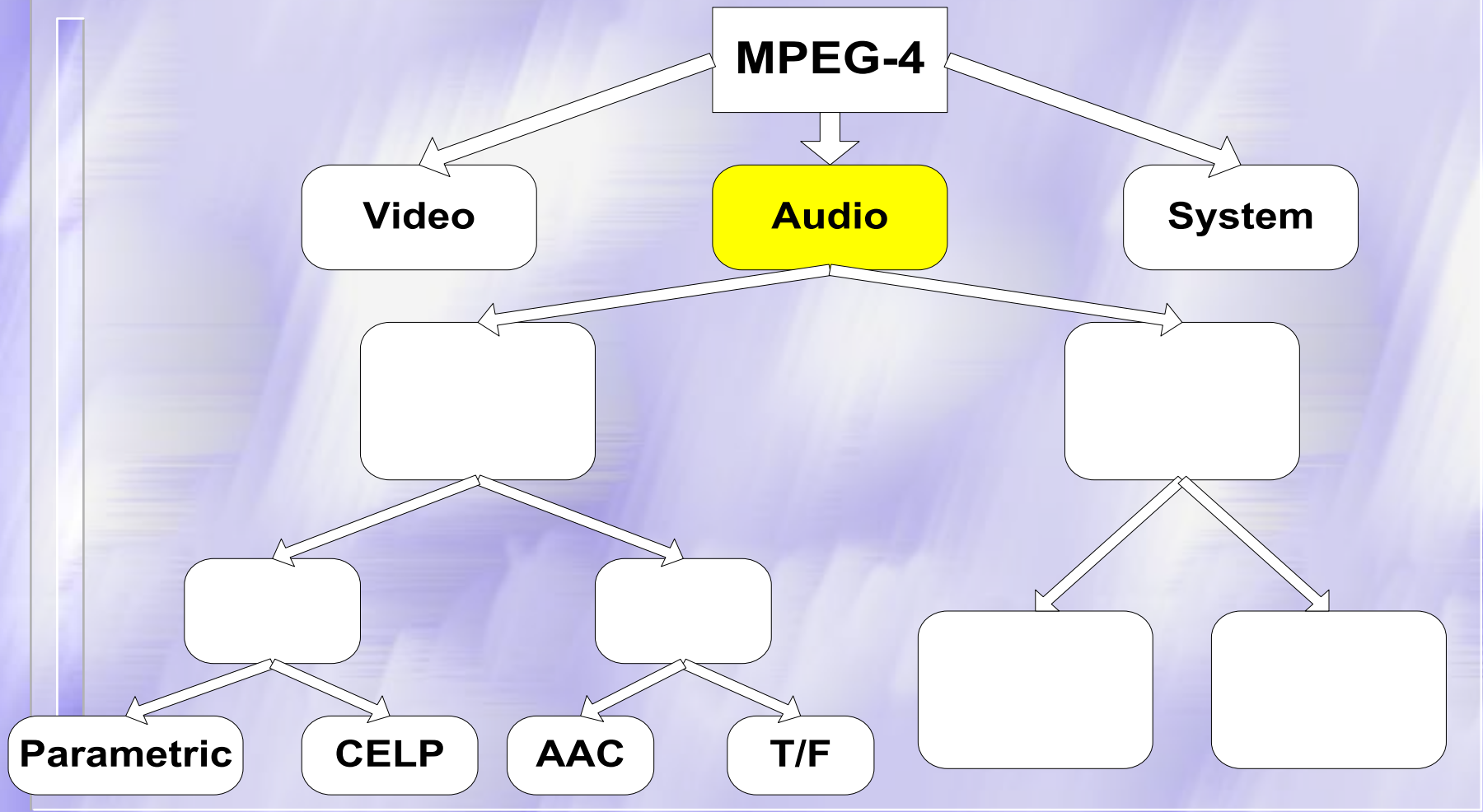
MPEG-4

GRAPHICS & MEDIA LAB
VIDEO GROUP

Разрабатывался как очередной стандарт в мире мультимедиа и его первый Profile был закончен в 1998.

В отличие от MPEG-1 и MPEG-2, в MPEG-4 акцент сделан в основном на функциональность, а не на повышение эффективности сжатия.

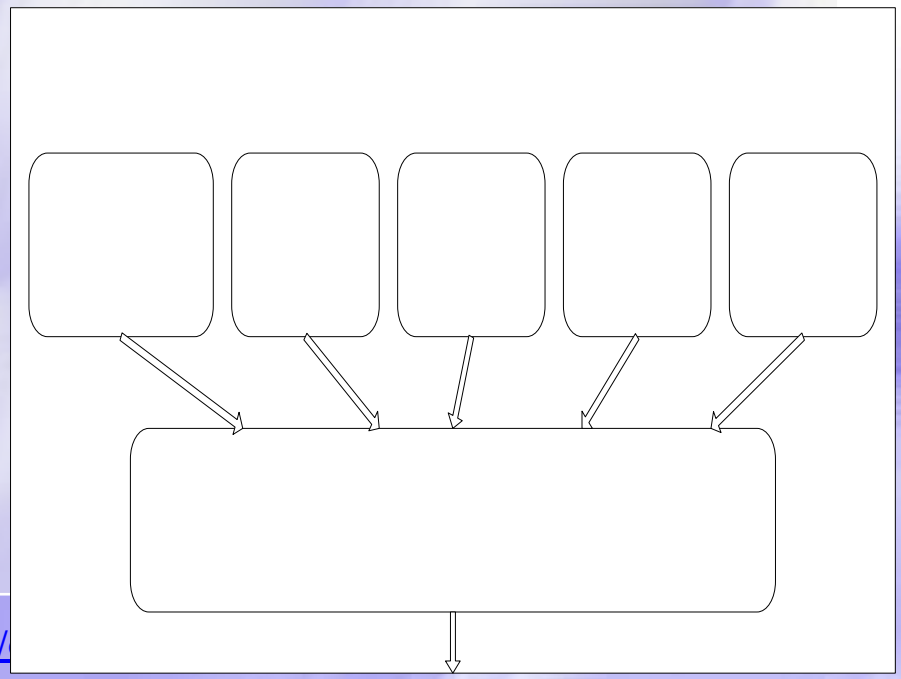
MPEG-4
Структура стандарта



Компоненты MPEG-4 Audio

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Содержит набор различных кодек для различных типов сигналов и диапазонов частот дискретизации
  - Parametric Speech и Audio Coder
  - CELP Speech Coder
  - General Audio (G/A) Coder
- Методы синтеза звука
  - Structure Audio System
  - Text to Speech Interface



MPEG-4 Audio

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Допустимы как synthetic coding и natural coding
- synthetic coding : вместо самого звука используется его описание. Приемник по описаниям создает похожий звук
- natural coding: Для звука используются 3 вида кодеров:
  - Параметрический кодер: для узкочастотной речи и звука частоты 2-4 кБт/с
  - CELP-кодер: для речи частоты 4-24 кбт/с
  - Перцепционный кодер: для звуковых сигналов частоты 4-24 кбт/с

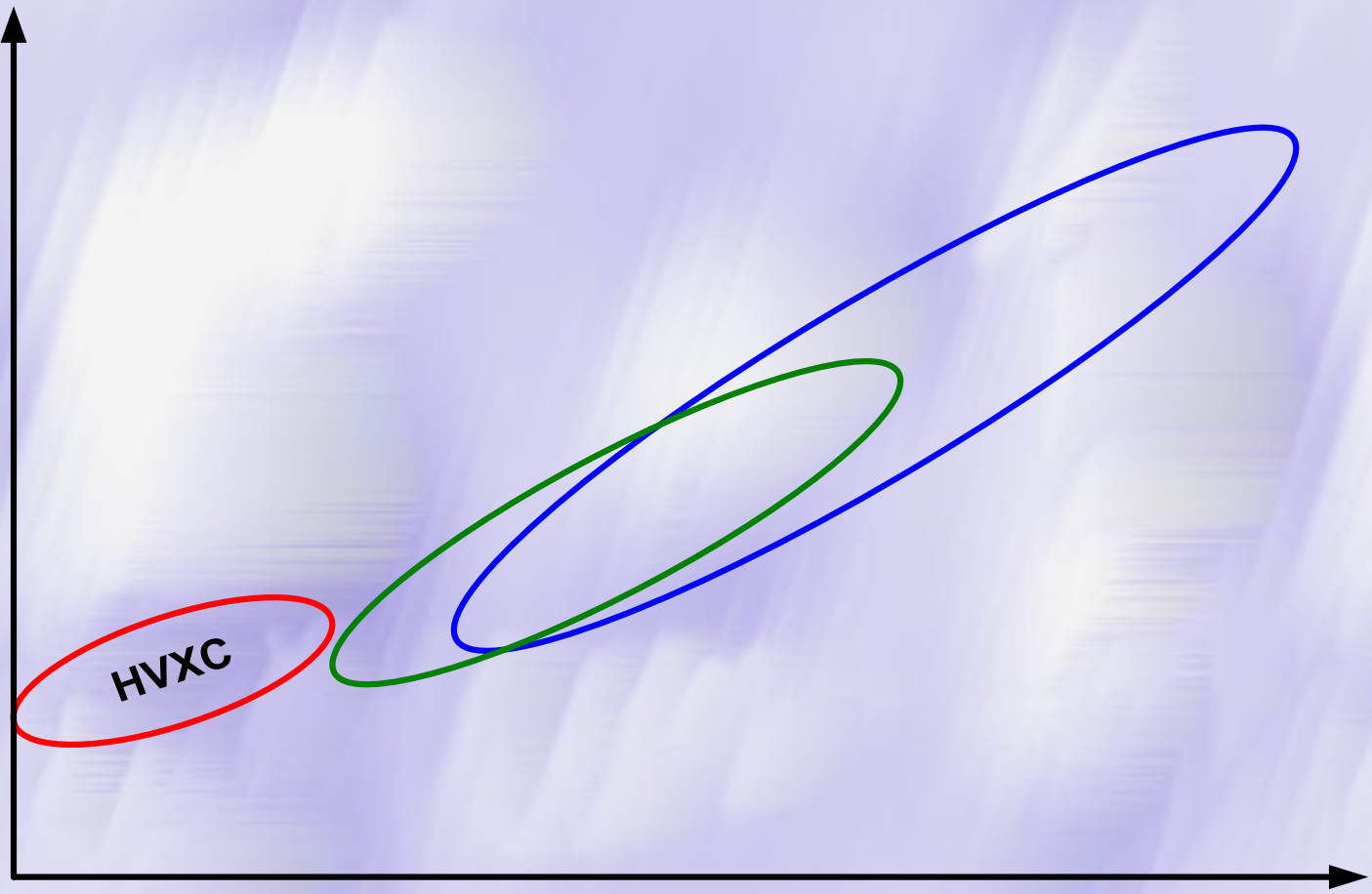
Обзор natural coding

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Различные средства, зависящие от битрейта и природы сигнала
- Средства могут комбинироваться (scalable coding)
- Широко применяются как в телефонных линиях, так и в высококачественных стереосистемах

MPEG-4
Natural audio

GRAPHICS & MEDIA LAB
VIDEO GROUP



MPEG-4
Кодирование речи

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Два основных алгоритма:
  - HVXC (Harmonic Vector eXcitation Coding)
  - CELP (Code Excited Linear Prediction)
- Широкая полоса битрейта: 1.5 – 24 кБит/с

MPEG-4
Кодирование речи

GRAPHICS & MEDIA LAB
VIDEO GROUP

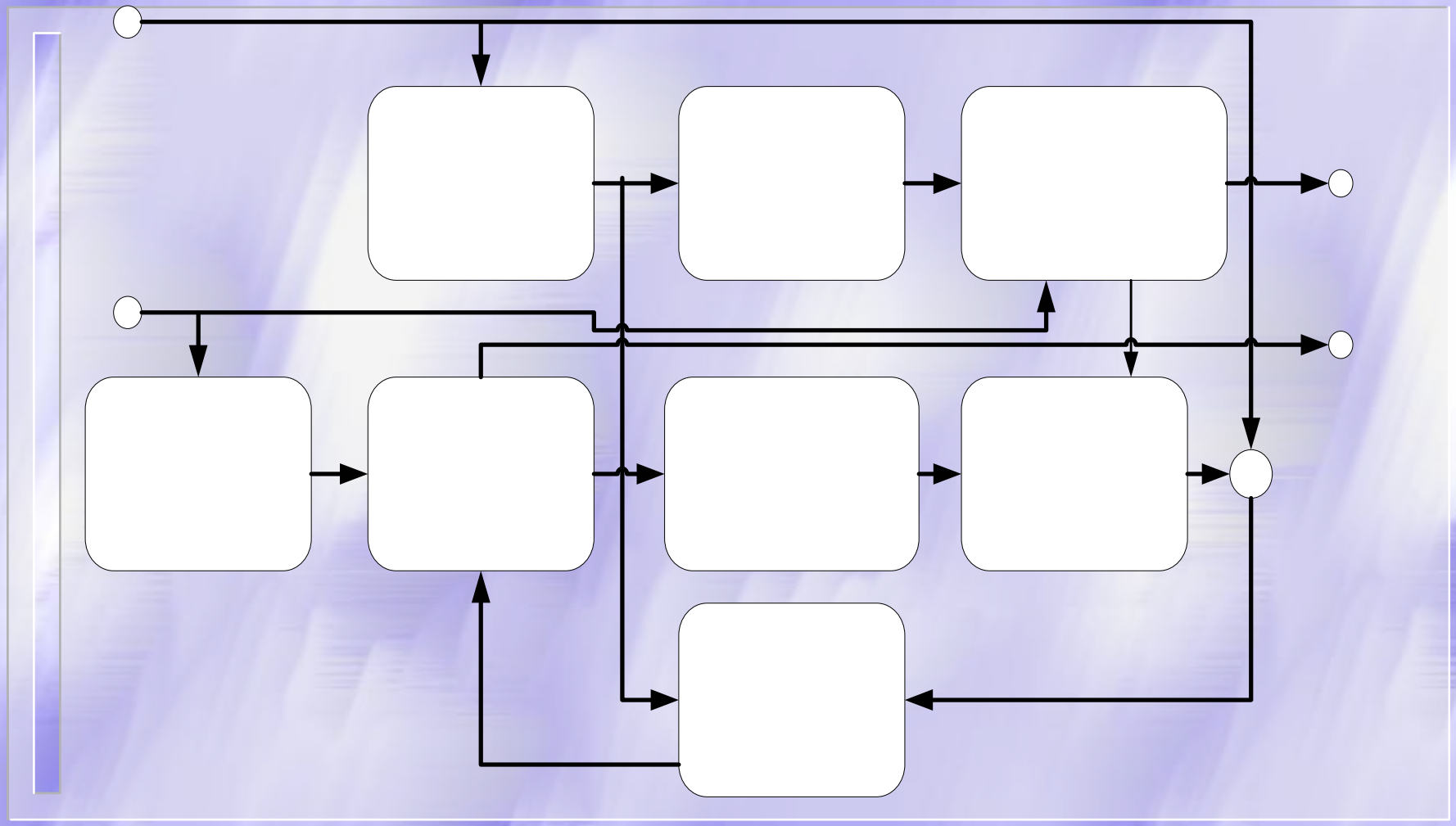
- CELP - для узкополосных и широкополосных каналов
- HVXC – как международный стандарт кодирования при самом низком битрейте (фикс. – 1.5 кБит/с и перем. – около 2.0 кБит/с)
- Новые возможности:
  - Скорость и изменение шага – HVXC
  - Регулирование битрейта – CELP, HVXC
  - Регулирование полосы частот – CELP

MPEG-4
CELP

GRAPHICS & MEDIA LAB
VIDEO GROUP

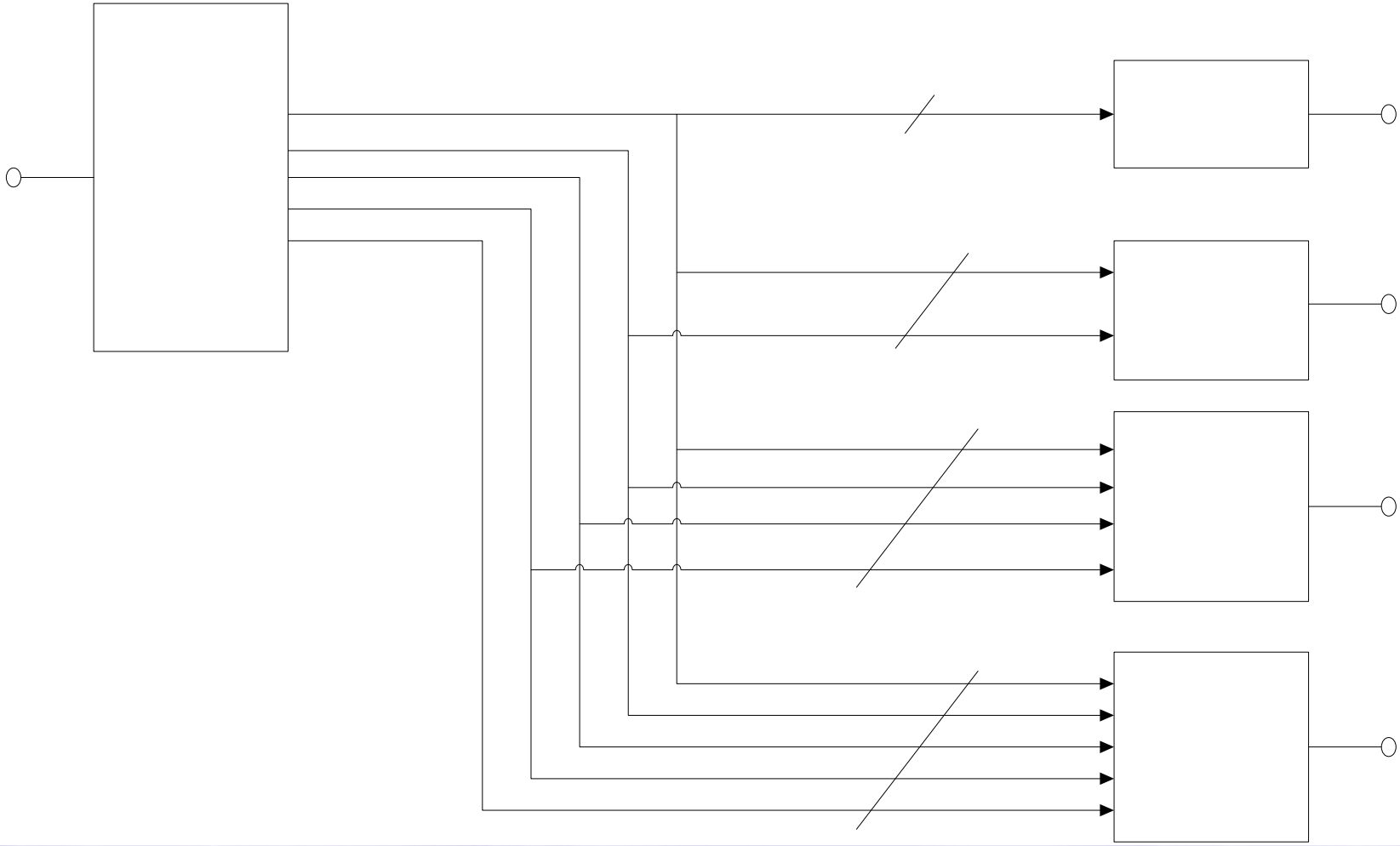
- Узкая полоса (NB): 3.85-12.2 кБит/с; 10-40 фреймов
- Широкая полоса (WB): 10.9-23.8 кБит/с; 10-20 фреймов
- Возможность менять шаг на 200-800 Бит/с
- Регулирование битрейта:
  - NB – шаг в 2.0 кБит/с
  - WB – шаг в 4.0 кБит/с
- Регулирование полосы частот
- Точное регулирование скорости
- Один импульс: WB – низкая сложность
- Много импульсов: WB, NB – высокая эффективность кодирования

Схема CELP-кодера



Входной речевой

Структура регулирования битрейта



MPEG-4
HVXC

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Низкий битрейт / хорошее качество
  - 2.0 / 4.0 кБит/с (фикс.); 1.5 / 3.0 кБит/с (перем.)
  - HVXC при 2.0 кБит/с имеет более высокое качество, чием FS1016 CELP при 4.8 кБит/с
- Регулирование битрейта
  - Декодирование при 2.0 кБит/с может использовать поток при 4.0 кБит/с
- Регулирование скорости и шага
  - Очень подходит для быстрого поиска в голосовой базе данных и для быстрых просмотров

MPEG-4
HVXC (подход)

GRAPHICS & MEDIA LAB
VIDEO GROUP

Объединены две схемы кодирования разных типов. Одна подходит для звучных участков. Другая – для глухих.

Звонкие участки – предсказанная волна вычитается из сигнала, а ошибка сжимается в частотную область
Глухие участки – обрабатываются кодером CELP

Обзор synthetic coding

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Вместо самого звука передает его параметрическое представление
- Допускает передачу со сверхнизкой полосой частот
- Музыка: Structured Audio (SA)
- Речь: Text-To-Speech interface

Structured Audio

GRAPHICS & MEDIA LAB
VIDEO GROUP

SA был изобретен компанией :

Machine Listening Group

Основная идея SA: передача звука осуществляется скорее по его описанию, чем с помощью его сжатия.

Structured Audio
(В чем проблема?)

GRAPHICS & MEDIA LAB
VIDEO GROUP

Существует ряд форматов сжатия аудио таких, как RealAudio, MP3, Liquid Audio, для передачи музыкальных файлов в интернете. Но у всех есть проблема: несоизмеримость качества звука с объемом музыкального файла.

Формат SA подразумевал приемлемое качество при достаточно небольшом объеме файла.

Structured Audio

GRAPHICS & MEDIA LAB
VIDEO GROUP

Принцип генерации аудио на ходу, который используется в SA, называют кодированием Колмогорова.

SA включает в себя мощный язык обработки звука SAOL (произносится «сэил») и язык оценки музыки SASL (произносится «сэссил»), с поддержкой существующего MIDI-формата.

Structured Audio

GRAPHICS & MEDIA LAB
VIDEO GROUP

Также в SA определено эффективное кодирование его элементов в удобный для хранения и передачи файл с двоичным форматом.

SA отличается от других форматов типа MIDI тем, что в нем задаются не только ноты, которые нужно проиграть, но и способы преобразования этих нот.

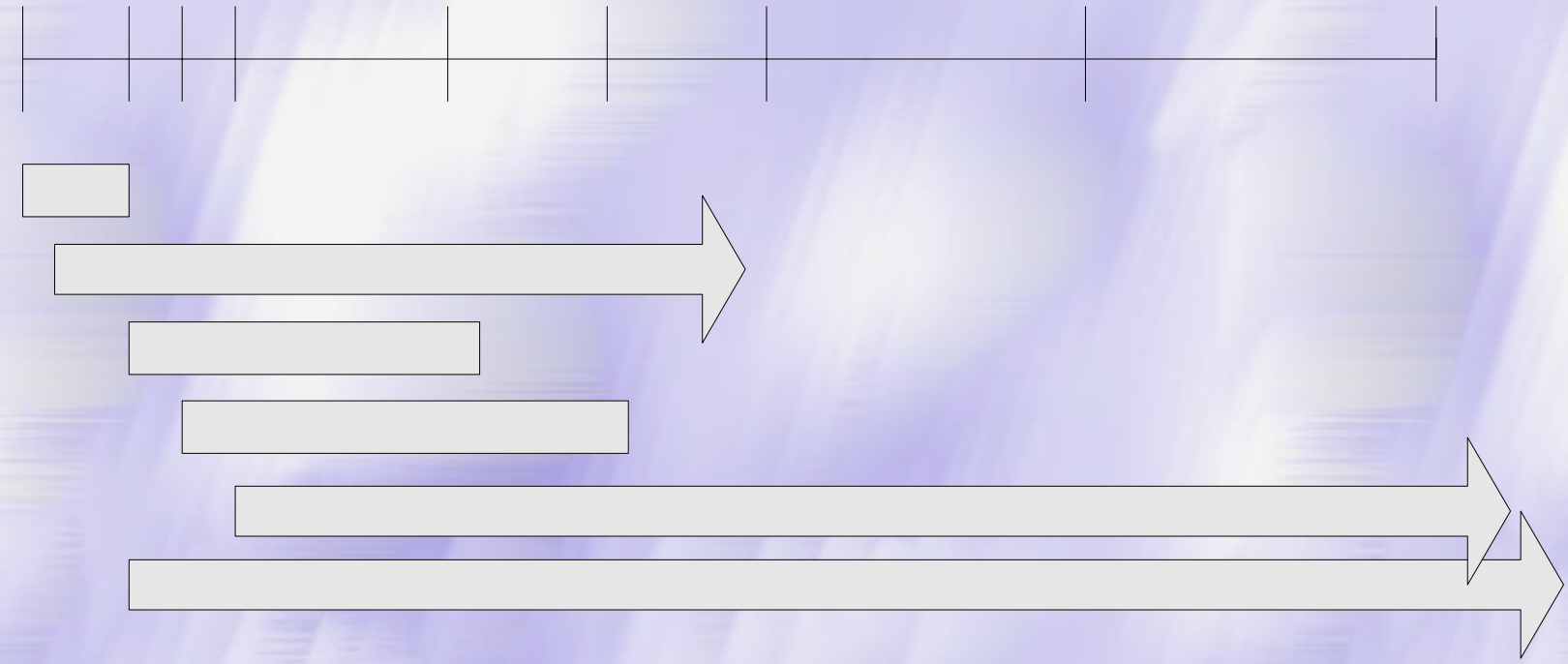
Structured Audio

GRAPHICS & MEDIA LAB
VIDEO GROUP

В то время, как инструментальные модели используют алгоритм синтеза вместо таблиц сигналов, файл SA может описать описать реалистичное музыкальное представление без использование аудиоданных.

Таким образов SA-файл звучит, как WAV, но имеет меньший в 50-1000 раз объем.

Параметры Audio-компонентов



MPEG-7

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Был утвержден в июле 2001 года.
- В отличие от MPEG-1/2/4, MPEG-7 не определял алгоритмов сжатия.
- MPEG-7 стал стандартом поиска, фильтрации, управления и обработки мультимедиа-информации.

Сжатие Dolby Audio

GRAPHICS & MEDIA LAB
VIDEO GROUP

Области применения:

- Спутниковое FM вещание, передача звука на TV (Dolby AC-1)
- Обычный стандарт кодирования в компьютерных звуковых картах (Dolby AC-2)
- Высококачественный стандарт ATV (advanced television). Область конкурирования MPEG (Dolby AC-3)

Отличия от MPEG

GRAPHICS & MEDIA LAB
VIDEO GROUP

- MPEG-кодеры контролируют точность квантования путем вычисления количества бит для каждого сэмпла.
- MPEG должен хранить каждое значение квантования вместе с каждым сэмплом
- MPEG-декодеры используют эту информацию для деквантования: forward adaptive bit allocation
- Преимущество MPEG состоит в том, что психоаккустическая модель не требуется в декодировании, где хранятся значения квантования

Отличия от MPEG

GRAPHICS & MEDIA LAB
VIDEO GROUP

DOLBY: используется фиксированное распределение битрейта.

- Не нужно посылать с каждым фреймом, как в MPEG
- Кодеры и декодеры DOLBY используют эту информацию Фиксированное распределение битрейта определяется исходя из свойств и характеристик чувствительности человеческого уха.

Различные стандарты
Dolby AC-1

- Простая психоаккустическая модель
- 40 частотных подполос в семплировании при 32 кБит/с
- Пропорционально большее число частотных подполос при 44.1 кБит/с и 48 кБит/с
- Обычное сжатие для 512 кБит/с для стерео

Различные стандарты
Dolby AC-2

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Возможность варьирования распределения битрейта
- Теперь декодер требует копии психоаккустической модели
- Encoded spectral envelope
- Backward adaptive bit allocation mode
- Высокое (hi-fi) качество звука при 256 кБит/с
- Не подходит для приложений вещания: кодер не может менять модель, не меняя декодера
- Обычное кодирование в компьютерных аудиокартах

Различные стандарты
Dolby AC-3

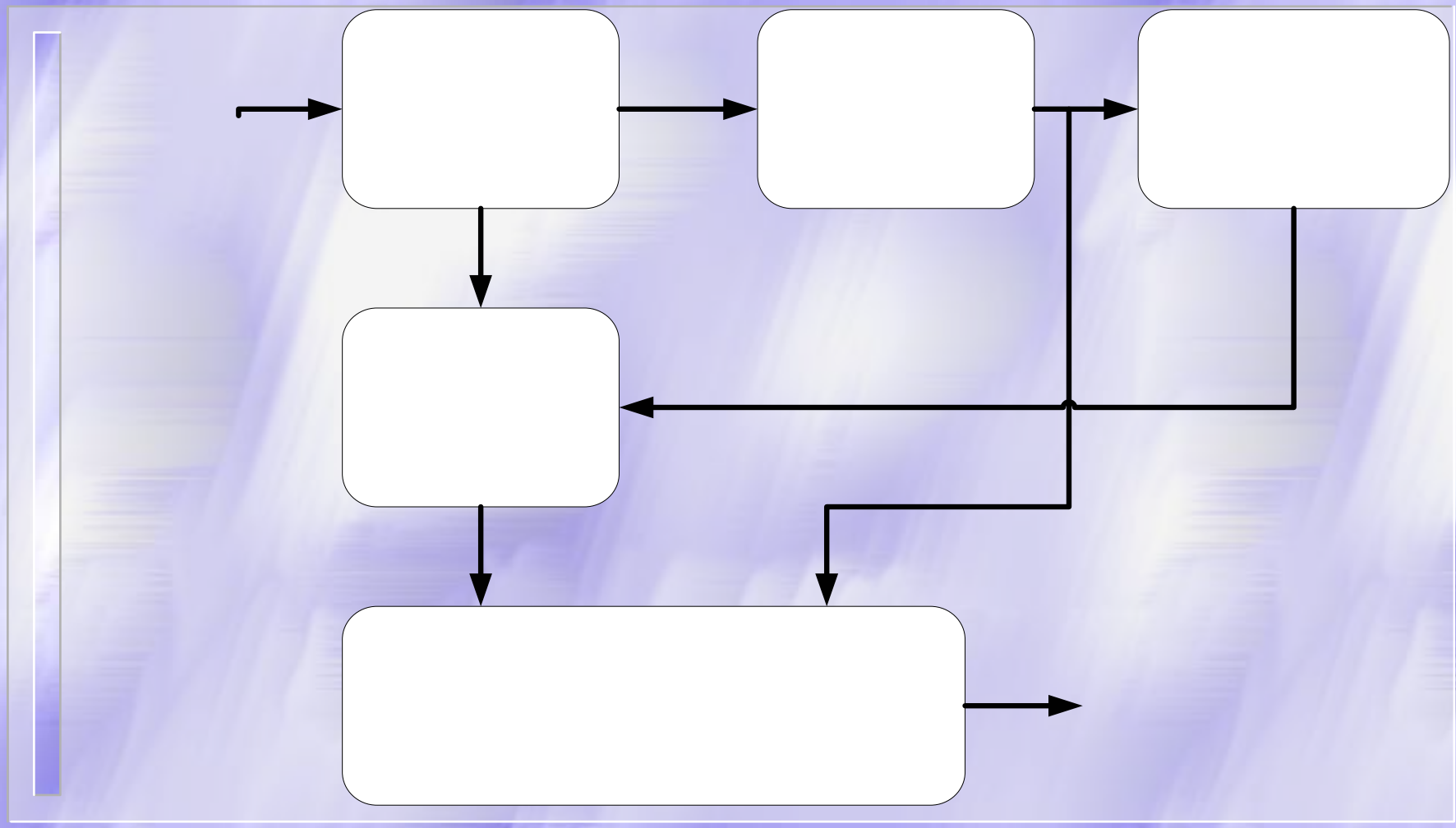
GRAPHICS & MEDIA LAB
VIDEO GROUP

Может кодировать от 1 до 5.1 каналов исходного звука из представления PCM в закодированный поток от 32 кБит/с до 640 кБит/с. Поддержка зависимости от ширины частотной полосы исходного сигнала

Использование смешанного режима: backward/forward adaptive bit allocation. Любая информация модификации модели кодируется во фрейме.

Используется в высокоточных ATV-стандартах. Алгоритм AC-3 достигает высокой степени сжатия путем грубого квантования представления частотной полосы аудиосигнала.

AC-3
(схема кодирования)



Временные

AC-3
кодирование

GRAPHICS & MEDIA LAB
VIDEO GROUP

Первый шаг в кодировании – это преобразование представления аудио из PCM семплов в последователь блоков частотных коэффициентов. Это происходит при анализе банком фильтров. Из перекрывающихся блоков по 512 семплов выделяется временное окно и переводится в частотную область.

AC-3
кодирование

GRAPHICS & MEDIA LAB
VIDEO GROUP

Из-за перекрытия блоков, каждый семпл представлен в двух последовательных преобразованных блоках. Представление частотной области может быть урезано до степени двойки, так что в каждом блоке будет содержаться 256 частотных коэффициентов.

AC-3 кодирование

GRAPHICS & MEDIA LAB
VIDEO GROUP

Каждый частотный коэффициент представляется в формат с плавающей точкой. Последовательности порядков образуют грубое представление спектра сигнала, которое называется спектральная огибающая. Она используется центральной процедурой размещения битов, которая определяет, сколько битов нужно использовать для кодирование каждой мантиссы.

AC-3
кодирование

GRAPHICS & MEDIA LAB
VIDEO GROUP

Спектральная огибающая и грубо квантованные мантиссы для 6 аудиоблоков (1536 семплов) кодируются в один AC-3 фрейм. Поток AC-3 представляет собой последовательность AC-3 фреймов.

AC-3
кодирование (чего нет в схеме)

GRAPHICS & MEDIA LAB
VIDEO GROUP

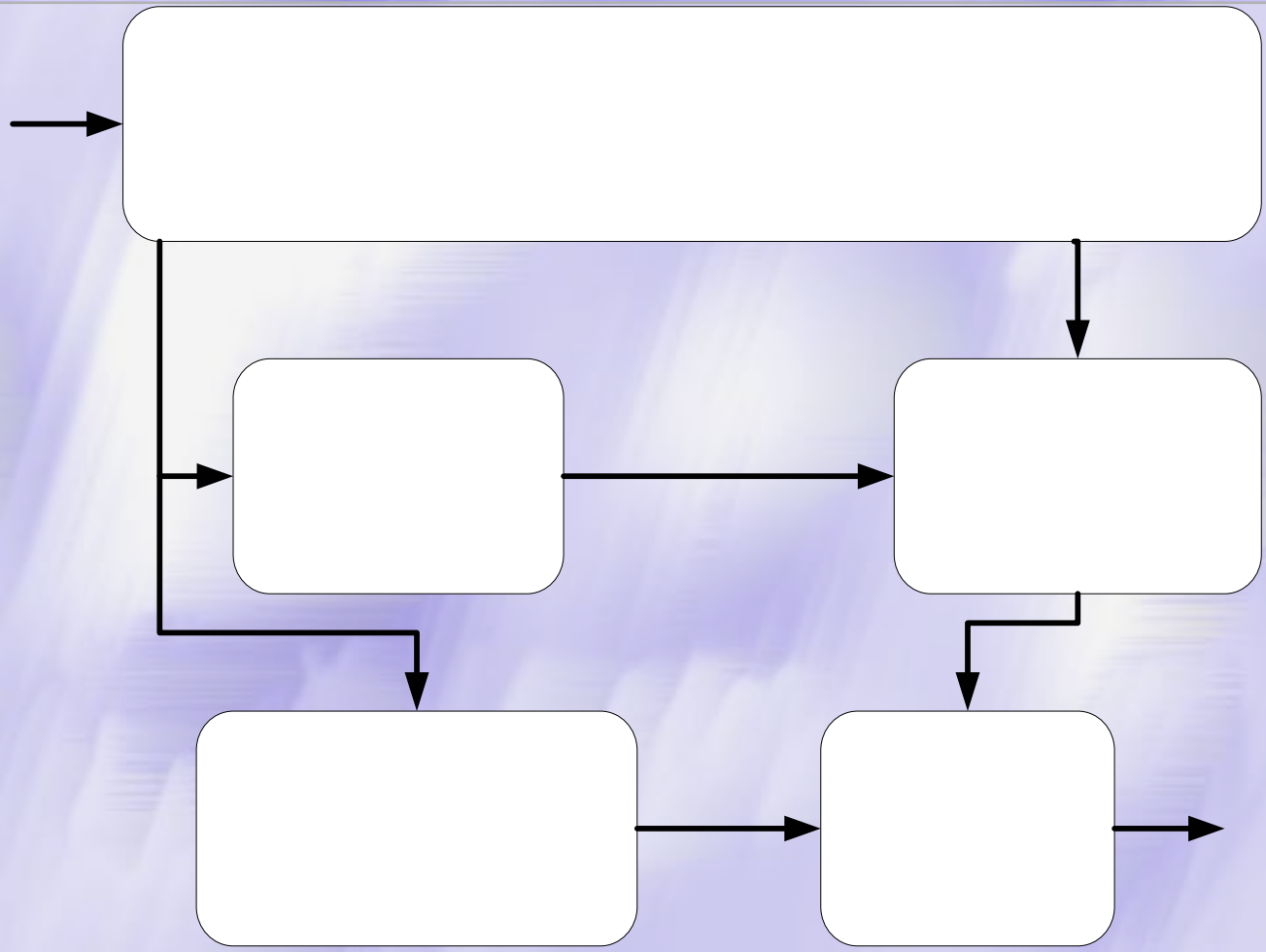
- Фреймовый заголовок
- Коды обнаружения ошибок
- Анализ банка фильтров
- Спектральная огибающая
- Размещение битов
- На высоких частотах каналы могут использовать общую информацию
- Matrixing (в двухканальном режиме)

AC-3 (декодирование)

GRAPHICS & MEDIA LAB
VIDEO GROUP

Процесс декодирования обратный к процессу кодирования. Декодер должен синхронизироваться по входному потоку, контролировать ошибки, преобразовывать разные типы данных, таких как закодированная спектральная огибающая и квантованные мантиссы. Из спектральной огибающей получаются порядки. Полученные плавающие числа преобразуются обратно во временную область.

AC-3
(схема декодирования)



Закодирова

AC-3
(чего нет в схеме декодирования)

GRAPHICS & MEDIA LAB
VIDEO GROUP

- «Приглушение» ошибок при их обнаружении
- Высокочастотные каналы, которые были склеены разъединяются
- Dematrixing
- Разрешение синтезированного банка фильтров должно динамически изменяться аналогично процессу кодирования

ГИБРИДНОЕ СЖАТИЕ АУДИО

Недостатки традиционной схемы кодирования

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Кодирование формы волны

  Сигналы с разной формой волны могут звучать одинаково:

  - Шумовые сигналы
  - Инвертированный сигнал
  - Смещенный сигнал
- Независимое кодирования фреймов
  - Музыка – совокупность повторяющихся видоизменяющихся звуков

Гибридный кодер: идеи

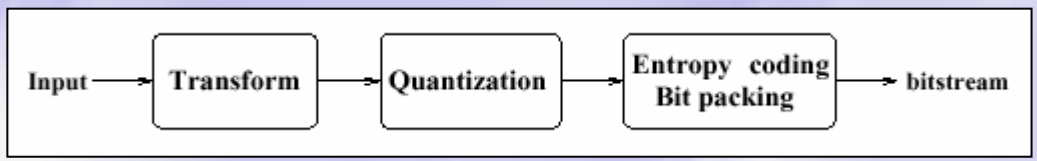
GRAPHICS & MEDIA LAB
VIDEO GROUP

- Разделение сигнала на три компоненты и использование разных типов кодирования в зависимости от специфики компоненты:
  - Гармоническую, используя преобразования, хорошо локализирующие частоты, основанные на Фурье
  - Ударные, используя Вейвлет преобразование, имеющее лучшую временную локализацию
  - Шумовую, используя кодирования энергетических огибающих спектра

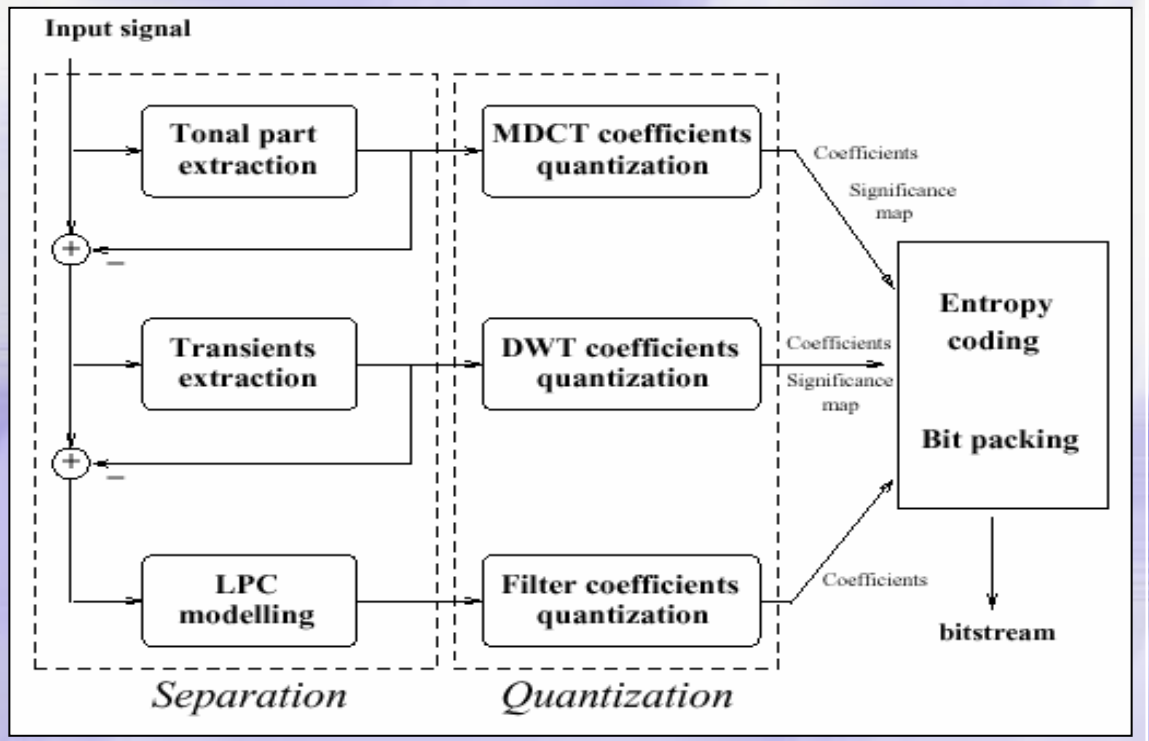
Схема гибридного и обычного аудио-кодера

GRAPHICS & MEDIA LAB
VIDEO GROUP

Обычный аудио-кодер



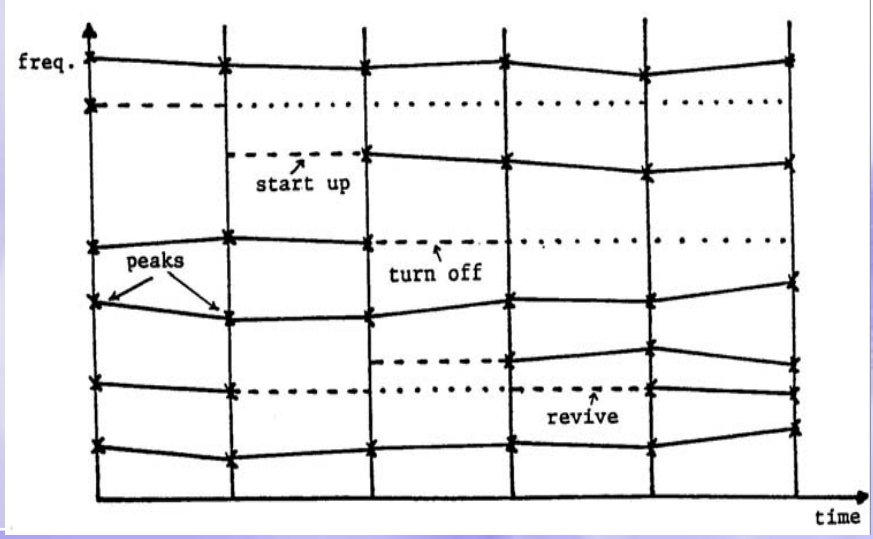
Пример гибридного кодека



Гибридный кодер: подходы к гармонической компоненте

GRAPHICS & MEDIA LAB
VIDEO GROUP

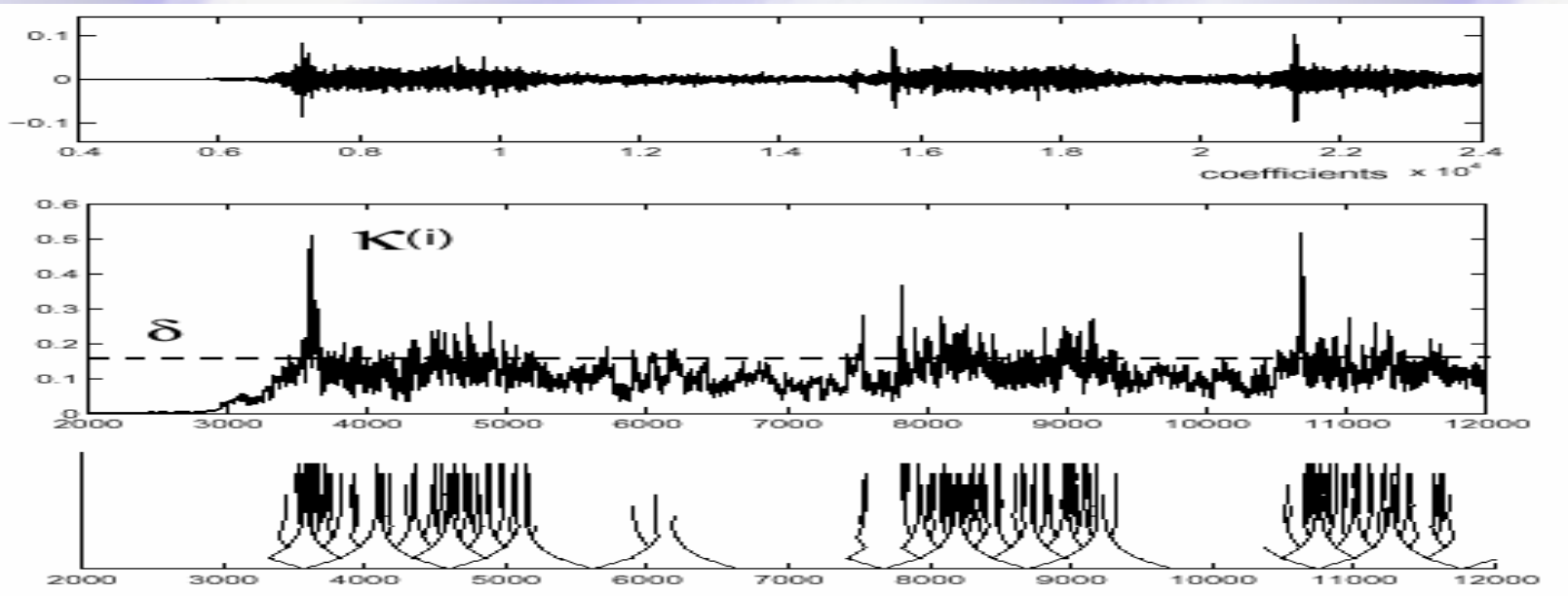
- Представление, основанное на MDCT-маске:
  - Гармоника –локально стационарная по времени выделяющаяся часть MDCT-квазиспектра
  - Обнуление не ‘гармонических’ коэффициентов
  - Традиционное сжатие гармонической части
- Векторное представление и сжатие гармоник:
  - При восстановлении используется интерполяция



Гибридный кодер: переходные сигналы (удары)

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Выделение высоко амплитудных выбросов в сигнале с удаленными гармониками
- Разложение и сжатие на основе одномерных ортогональных вейвлетов



Гибридный кодер: представление остатка

GRAPHICS & MEDIA LAB
VIDEO GROUP

Остаток = Сигнал – Гармоники– Переходные

- Гипотеза: остаток = шумовой сигнал
- Для кодирования шума используются LPC кодирование спектральной огибающей
- Для реконструкции используется фильтрация белого шума

Компоненты: формы сигнал

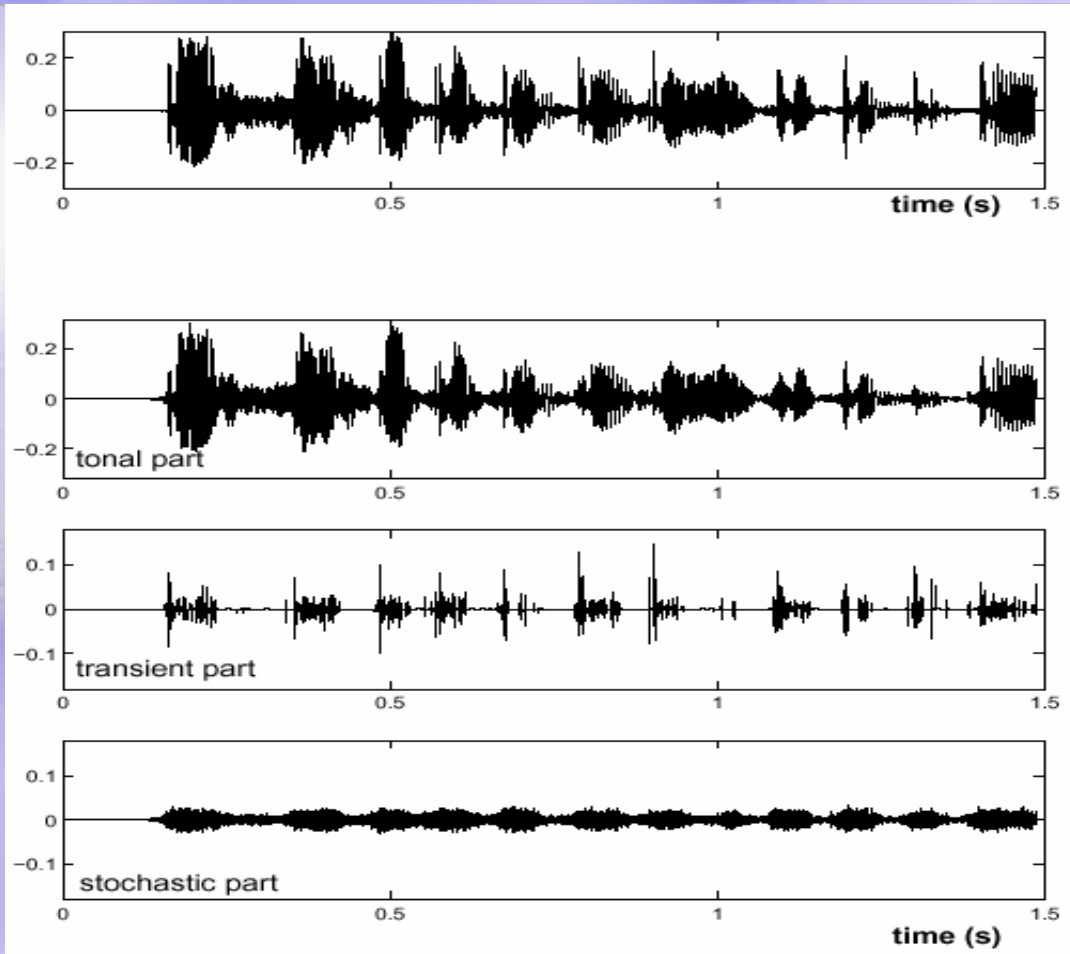
GRAPHICS & MEDIA LAB
VIDEO GROUP

Сигнал

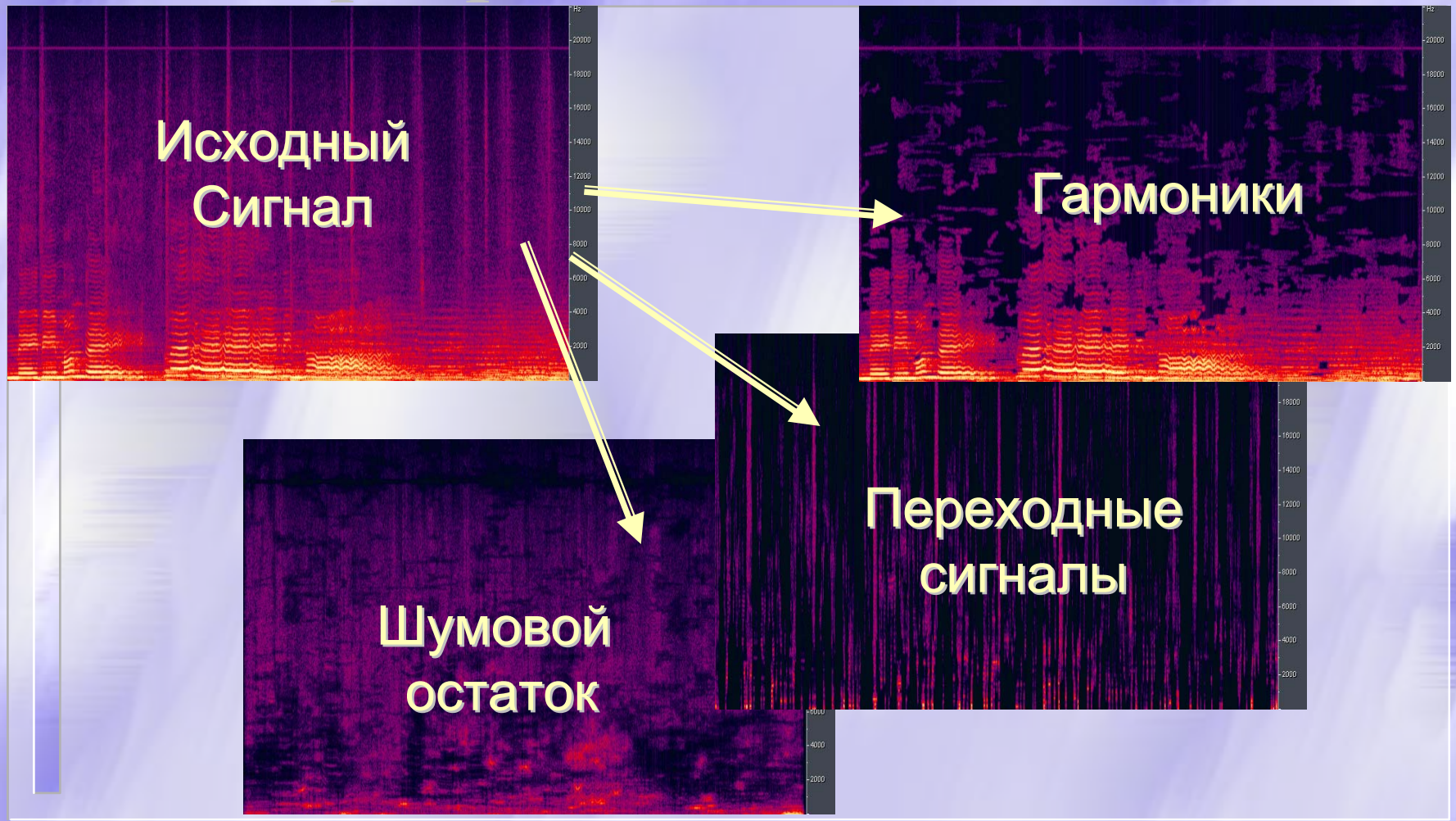
Гармоники

Переходная компонента

Шумовой остаток



Компоненты: спектрограммы



Гибридное кодирование: Выводы

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Преимущества:
  - Адаптивное кодирование, зависящее от конкретных свойств компонент
- Недостатки:
  - Избыточное представление
  - Аддитивный синтез шума – не устойчивость при итерационном применении
  - Ориентация на большое сжатие, но в настоящее время, абсолютно прозрачное кодирования с CD-качеством не достигнуто

СЖАТИЕ РЕЧИ

Специфика
Физическая и математическая модели
Пример кодека

Сжатие речи: Специфика

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Музыка:
  - Стандартный формат для представления музыки **Stereo 16 bit 44KHz** позволяет передавать весь диапазон слышимых человеком частот ~ 20Гц-20КГц
  - Несжатый поток: 1408 kb
  - Прозрачное кодирование (MP3): 128 kb (~10 раз)
- Речь:
  - Узкий частотный диапазон, реально от 70 до 3КГц, для передачи информативной части голоса достаточно: **Mono 8 bit 8KHz**
  - Несжатый поток: 64 kb
  - Прозрачное кодирование (GSM 6.1): 8 kb (8 раз)

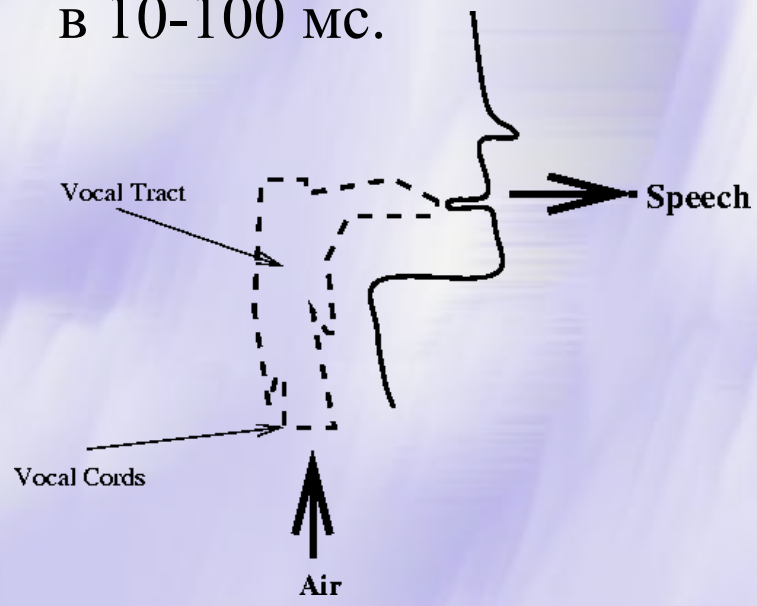
Критерии кодирования речи

GRAPHICS & MEDIA LAB
VIDEO GROUP

- Скорость передачи данных:
  - Фонематическая скорость: примерно 50 б/с
  - Познавательная скорость: примерно 400 б/с
  - Как к этим скоростям приблизиться?
- Понятность
- Естественность, качественность
- Вычислительная сложность
- Сложность реализации
- Максимальное время между получением замера и выходом закодированного значения
- Устойчивость к ошибкам

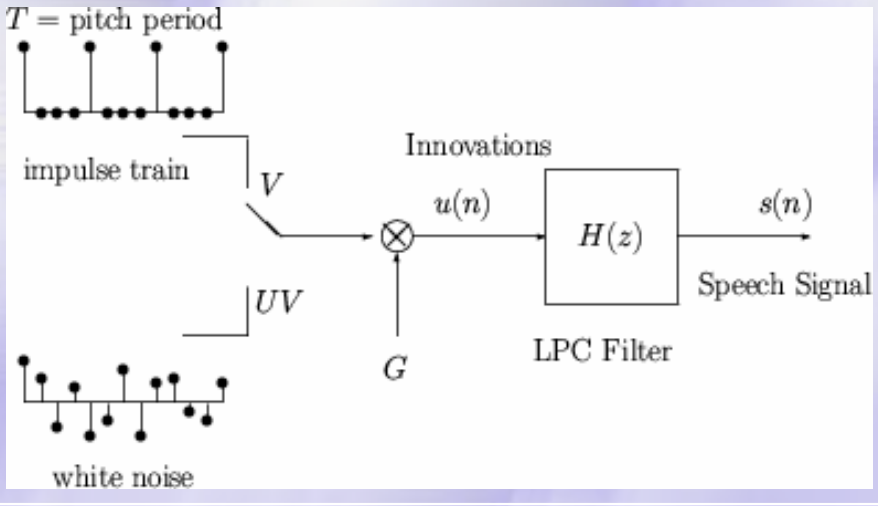
Физическая модель речи

- Гласные звуки заставляют вибрировать звуковой тракт. Скорость вибрации определяет *основной тон* голоса. Женщины и дети имеют высокий основной тон, мужчины низкий.
- Согласные оставляют голосовые связки стационарно открытыми
- При смене звука меняется форма речевого тракта. Смена происходит обычно раз в 10-100 мс.



Математическая модель речи

- Сигнал разбивается на фреймы, внутри которых считается что физическая модель постоянна
- Продолжительность фрейма обычно 20 мс, что соответствует 160 сэмплам



$$H(z) = \frac{1}{1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_{10} z^{-10}}$$

$$s(n) + \sum_{i=1}^{10} a_i s(n-i) = u(n)$$

13 параметров/фрейм:

$$\mathbf{A} = (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, G, V/UV, T)$$

Сжатие речи: пример 2.4 kb LPC Vocoder

GRAPHICS & MEDIA LAB VIDEO GROUP

- Для повышения устойчивости к квантованию, вместо LPC используют LSP(line spectrum pair), получающиеся преобразованием LPC
- Распределение бит:

Parameter Name	Parameter Notation	Rate (bits/frame)
LPC (LSP)	$\{a_k\}_{k=1}^{10}$ ($\{\omega_k\}_{k=1}^{10}$)	34
Gain	G	7
Voiced/Unvoiced & Period	$V/UV, T$	7
Total		48

LSP	No. of Bits
ω_1	3
ω_2	4
ω_3	4
ω_4	4
ω_5	4
ω_6	3
ω_7	3
ω_8	3
ω_9	3
ω_{10}	3
Total	34

V/UV	T	Encoded Value
UV	—	0
V	20	1
V	21	2
V	22	3
V	23	4
⋮	⋮	⋮
⋮	⋮	⋮
V	146	127