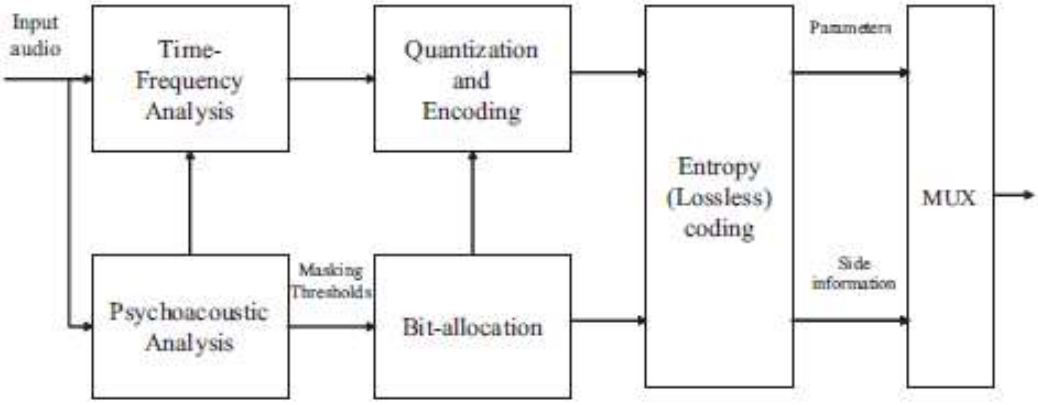
Сжатие аудио

Общая схема аудиокодеров



Параметры звукового сигнала

Интенсивность звука $I = \dfrac{P^2}{Q_0}, Вт/м^2$, $\quad Q_0 = 1{,}23\,кГ/м^3$ (1)

слуховое ощущение $E = a \lg I + C$ (2)

На пороге слышимости $I = I_{ПС}$

$$C = -a \lg I_{ПС}, \quad (3)$$

$$E = a \lg(I/I_{ПС}). \quad (4)$$

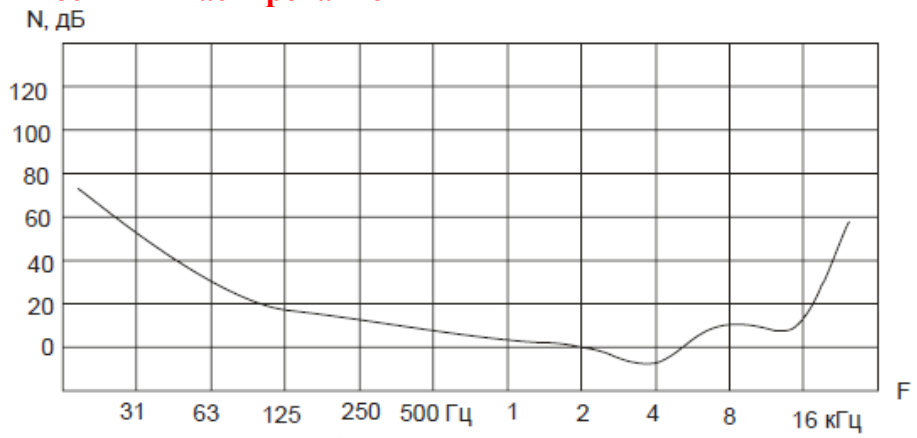
"бел" $(a = 1)$ "децибел" $(a = 10)$ $(I_0 = 10^{-12}\ Вт/м^2,\ P_0 = 2{,}04 \cdot 10^{-5}\ Па)$

уровень интенсивности $N(I) = 10 \lg(I/I_0), дБ$. (5)

уровень звукового давления $N = 20 \lg(P/P_0)$ (6)

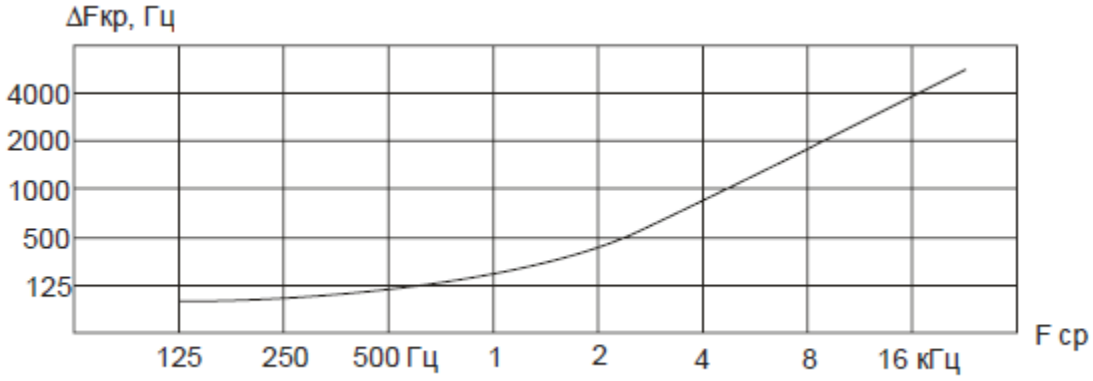
давление Н/м^2	дБ SPL	источник
2000.0	160	реактивный самолет
200.0	140	болевой порог
20.0	120	заводской шум
2.0	100	уличный транспорт
0.2	80	бытовой прибор
0.02	60	беседа
0.002	40	тихая комната
0.0002	20	шепот
0.00002	0	порог слышимости

Порог слышимости и Маскирование



Абсолютный порог слышимости

нижние границы критических полос слуха: 20, 100, 300, 400, 510, 630, 770, 920, 1080, 1270, 1480, 1720, 2000, 2320, 2700, 3150, 3700, 4440, 5300, 6400, 7700, 9500, 12000, 16000 Гц.



Зависимость ширины критической полосы слуха от ее средней частоты

$$H_z(f) = 1.3 \arctan(0.00076 f) + 3.5 \arctan\left[\left(\frac{f}{7500}\right)^2\right] \text{ (Bark)}$$

$$1 \text{ барк} = \begin{cases} \frac{f}{100} & \text{для частот } f < 500 \text{ Гц}, \\ 9 + 4\log\left(\frac{f}{1000}\right) & \text{для частот } f \geq 500 \text{ Гц}. \end{cases}$$

Band No.	Bandwidth (Hz)	Band No.	Bandwidth (Hz)
1	- 100	14	2000-2320
2	100-200	15	2320-2700
3	200-300	16	2700-3150
4	300-400	17	3150-3700
5	400-510	18	3700-4400
6	510-630	19	4400-5300
7	630-770	20	5300-6400
8	770-920	21	6400-7700
9	920-1080	22	7700-9500
10	1080-1270	23	9500-12000
11	1270-1480	24	12000-15500
12	1480-1720	25	15500-25000
13	1720-2000	26	25000-

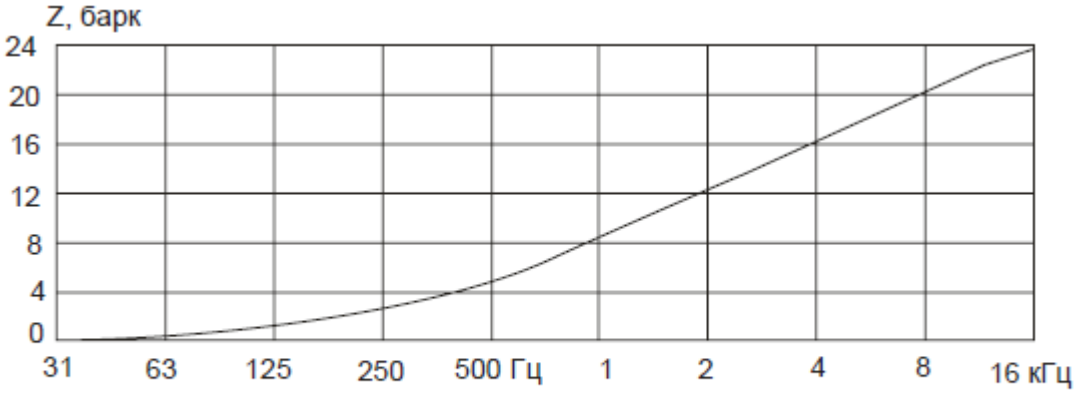
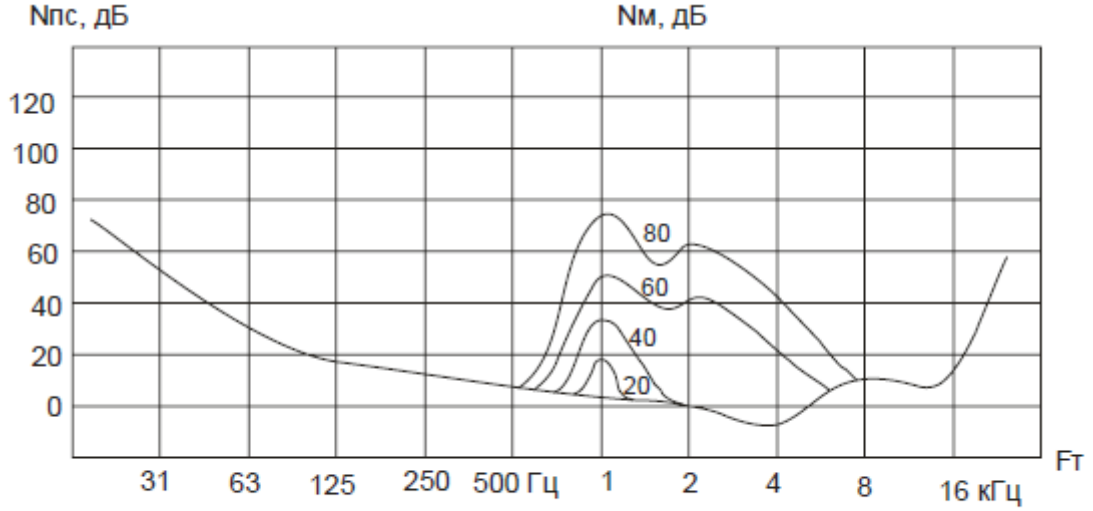
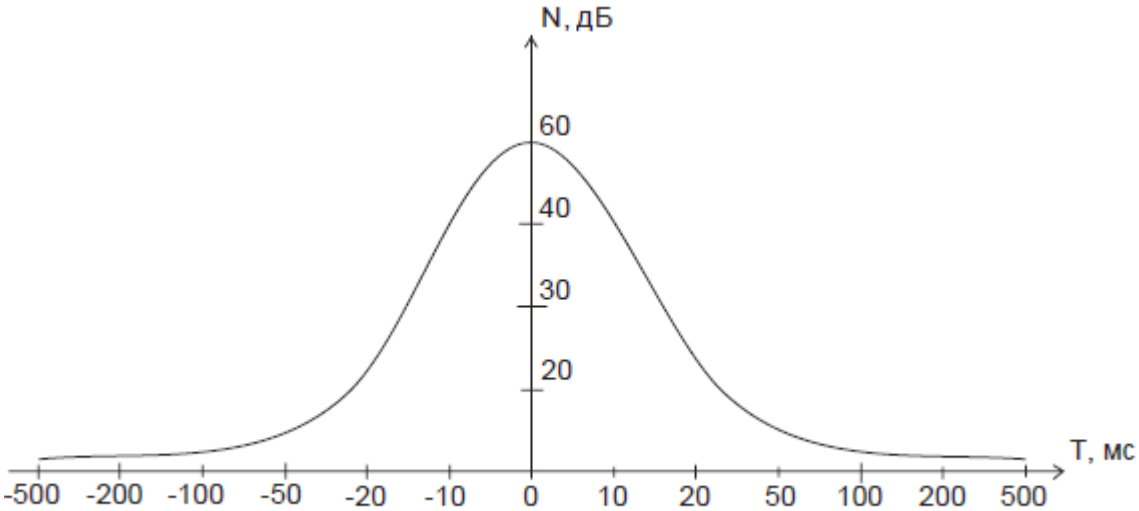


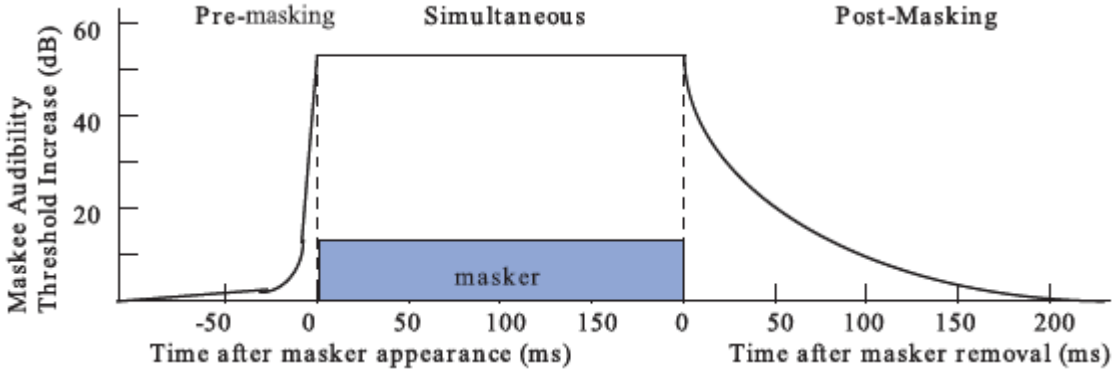
График перехода от частотной шкалы высоты тона в октавах к шкале высоты тона в барках



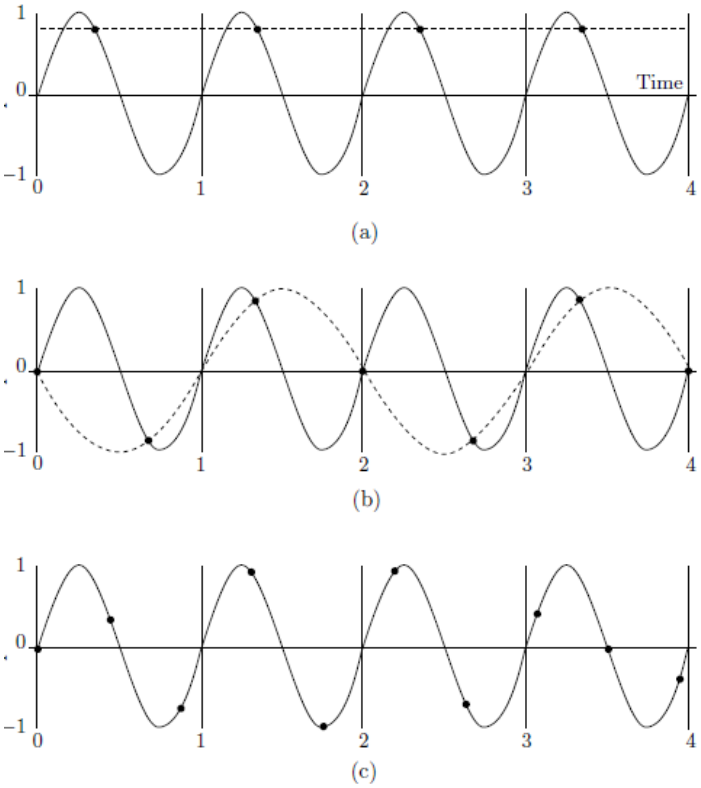
Кривые порога слышимости тона с частотой F_T при маскировке тоном с частотой F_M=1 кГц и уровнем N_M, дБ



Порог слышимости при временном маскировании



Семплирование



Wave audio format

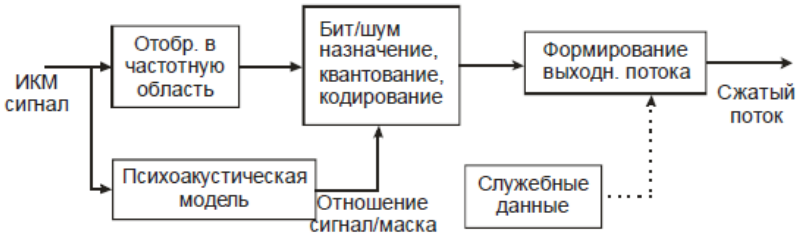
Byte #	Description
0–3	fmt␣
4–7	Length of subchunk (24+extra format bytes)
8–9	16-bit compression code
10–11	Channel numbers (1=Mono, 2=Stereo)
12-15	Sample Rate (Binary, in Hz)
16–19	Bytes Per Second
20–21	Bytes Per Sample: 1=8 bit Mono, 2=8 bit Stereo or 16 bit Mono, 4=16 bit Stereo
22-23	Bits Per Sample
24–	Extra format bytes

Code	(Hex)	Compression
0	0000	Unknown
1	0001	PCM/uncompressed
2	0002	Microsoft ADPCM
6	0006	ITU G.711 A-law
7	0007	ITU G.711 μ-law
17	0011	IMA ADPCM
20	0016	ITU G.723 ADPCM (Yamaha)
49	0031	GSM 6.10
64	0040	ITU G.721 ADPCM
80	0050	MPEG
65,536	FFFF	Experimental

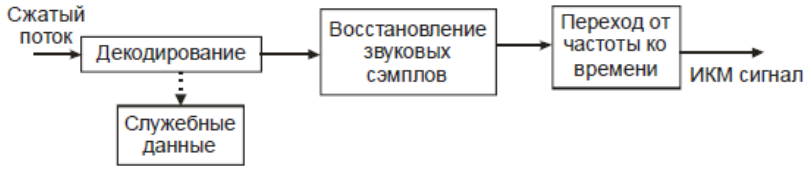
```
                    Hexadecimal                                  ASCII
00   52 49 46 46 E6 16 00 00 57 41 56 45 66 6D 74 20   RIFF....WAVEfmt␣
10   10 00 00 00 01 00 01 00 10 27 00 00 20 4E 00 00   .........'...N..
20   02 00 10 00 64 61 74 61 C2 16 00 00 8B FC 12 F8   ....data........
30   B9 F8 EB F8 25 F8 31 F3 67 EE 66 ED 80 EB F1 E7   ....%.1.g.f.....
40   3D EB 5C 09 C3 1B 44 E9 EF FF 18 39 35 0E 5B 0A   =.\...D....95.[.
```

Общая схема аудиокодера и аудиодекодера MPEG-1

Кодер MPEG.



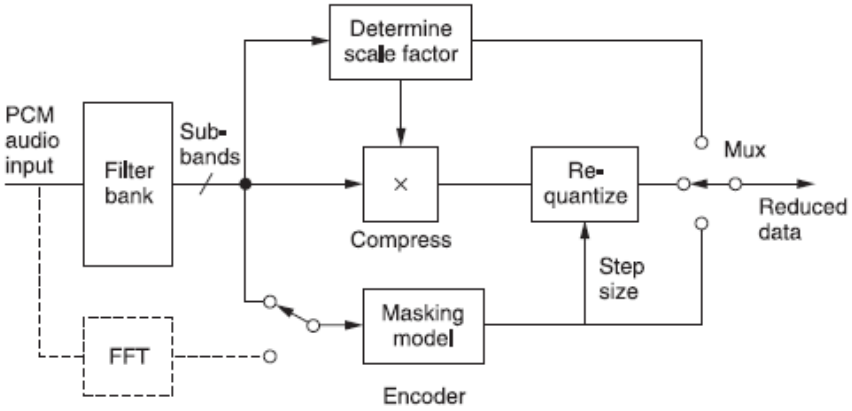
декодер MPEG.

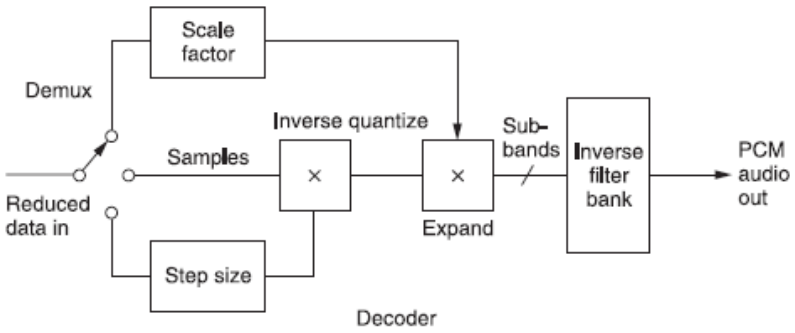


Bitrates in the Three Layers

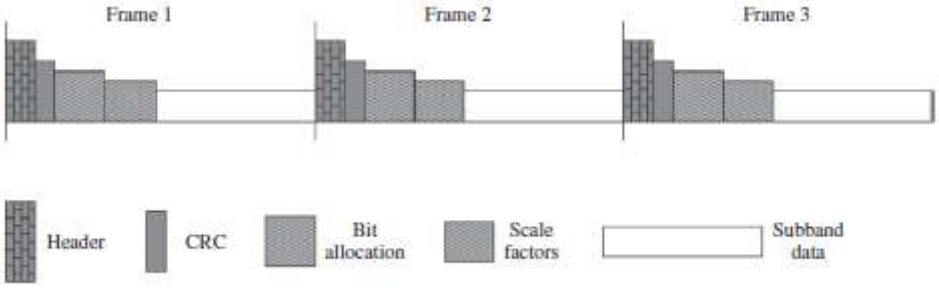
Index	Bitrate (Kbps) Layer I	Layer II	Layer III
0000	free format	free format	free format
0001	32	32	32
0010	64	48	40
0011	96	56	48
0100	128	64	56
0101	160	80	64
0110	192	96	80
0111	224	112	96
1000	256	128	112
1001	288	160	128
1010	320	192	160
1011	352	224	192
1100	384	256	224
1101	416	320	256
1110	448	384	320
1111	forbidden	forbidden	forbidden

MPEG Layer I





Frame structure for Layer 1



Формат заголовка

Поле 1. Идентификационный бит ID MPEG, значение равно 1

Поле 2. 11 – слой I, 10 – слой II и 01 – слой III. 00 зарезервировано.

Поле 3. Бит индексации использования защиты от ошибок.

Поле 4. Четыре бита для обозначения битовой скорости.

Поле 5. Два бита для скоростей сэмплирования.
00 — 44.1 кГц, 01 — 48 кГц и 10 — 32 кГц. 11 зарезервировано.

Index	MPEG-1	MPEG-2
00	44.1 kHz	22.05 kHz
01	48 kHz	24 kHz
10	32 kHz	16 kHz
11	Reserved	

Поле 6. Один бит, указывающий на использование заполнения.

$$\sum_{\text{первый кадр}}^{\text{последний кадр}} \frac{\text{размер кадра} \times \text{битовая скорость}}{\text{частота сэмплирования}},$$

размер кадра равен 384 для слоя I и 1152 для слоев II и III.

Поле 7. Один бит для частного использования кодером.

Поле 8. Двухбитное поле для указания стереозвука. Значение 00 – стереозвук, 01 – объединенное стерео (интенсивность–стерео или ms-стерео), 10 – двойной канал, 11 – одиночный канал.
среднее значение M_i и краевое значение S_i вместо левого и правого каналов L_i и R_i. Значения середина-край вычисляются по формулам

$$L_i = \frac{M_i + S_i}{\sqrt{2}} \text{ и } R_i = \frac{M_i - S_i}{\sqrt{2}}.$$

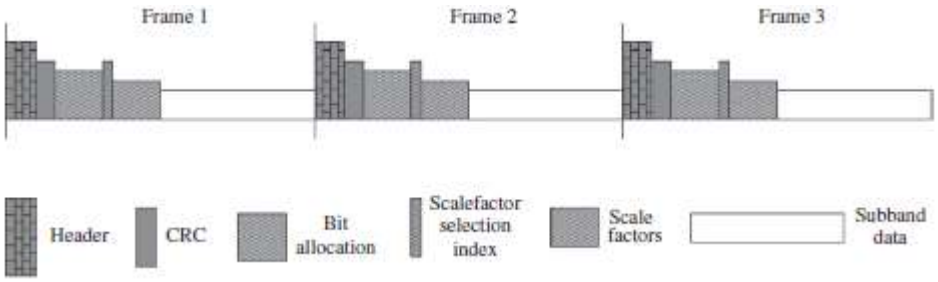
Поле 9. Двухбитовое поле расширения.

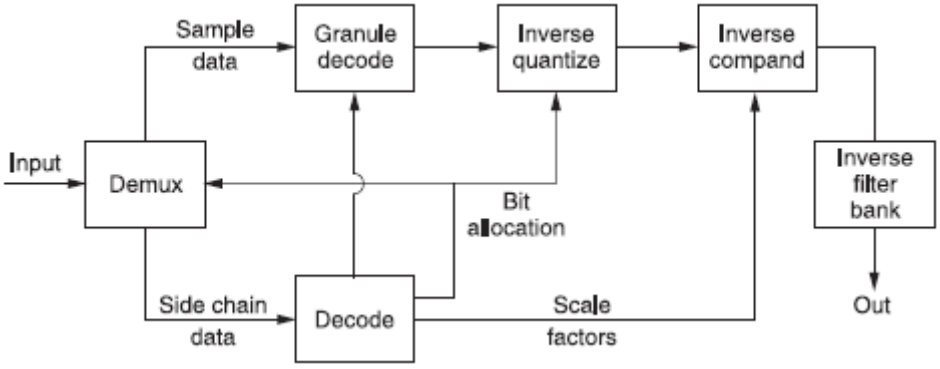
Поле 10. Бит копирайта.

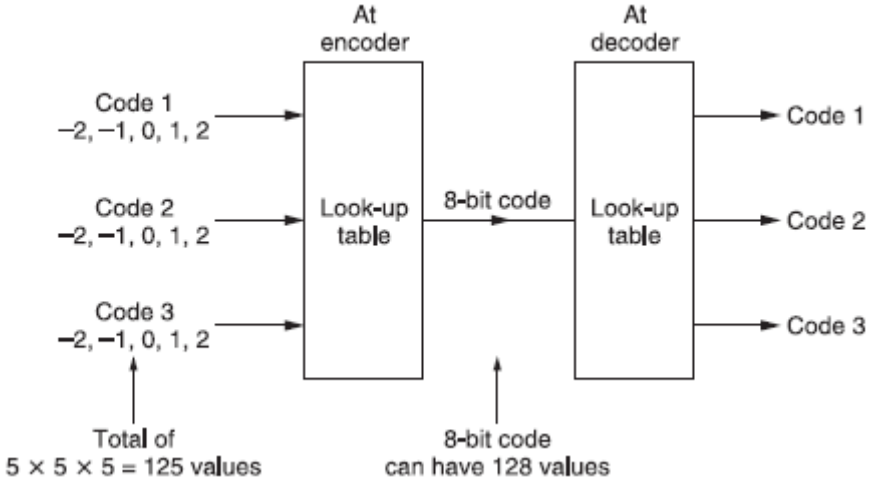
Поле 11. Один бит для обозначения оригинала или копии.

Поле 12. Двухбитовое поле усиления.

MPEG Layer II

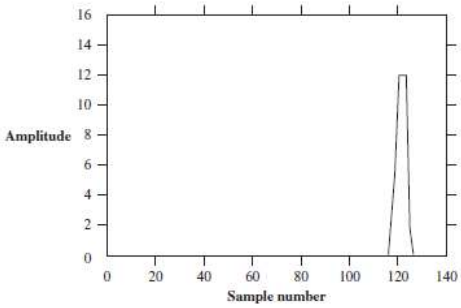


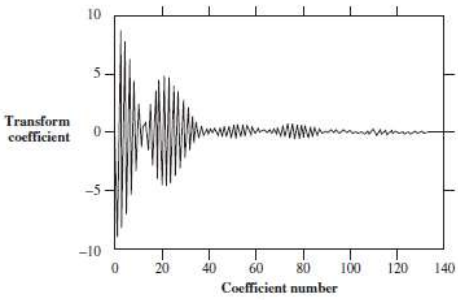




MPEG Layer 3 (MP3)

Появление пред-эха





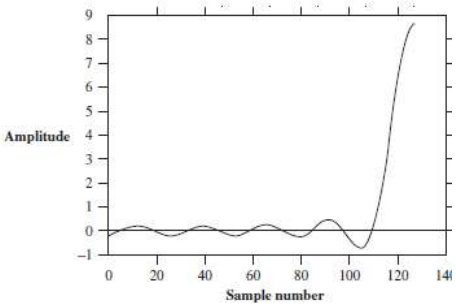
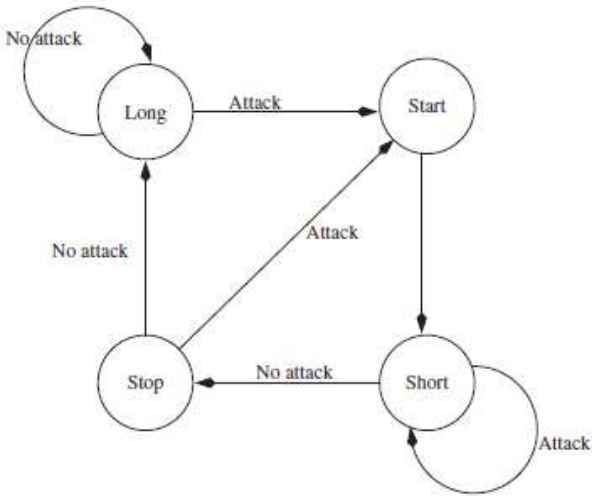
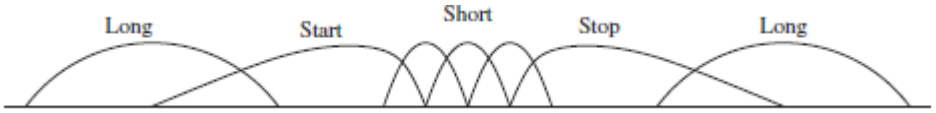
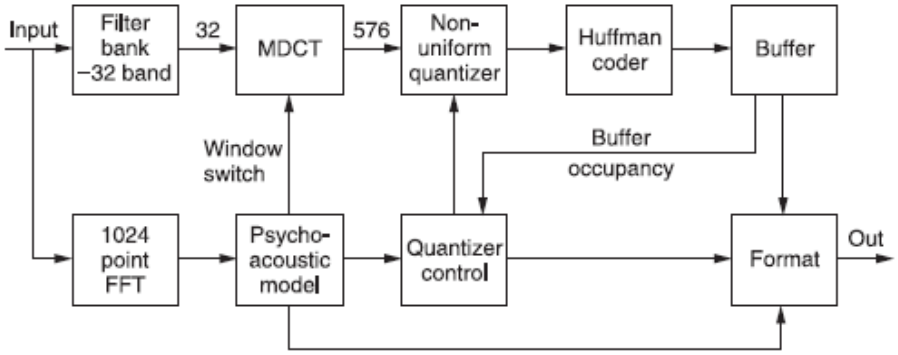


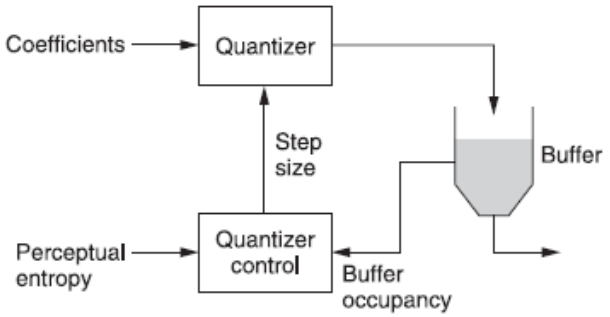
Диаграмма состояний для процесса переключения окон



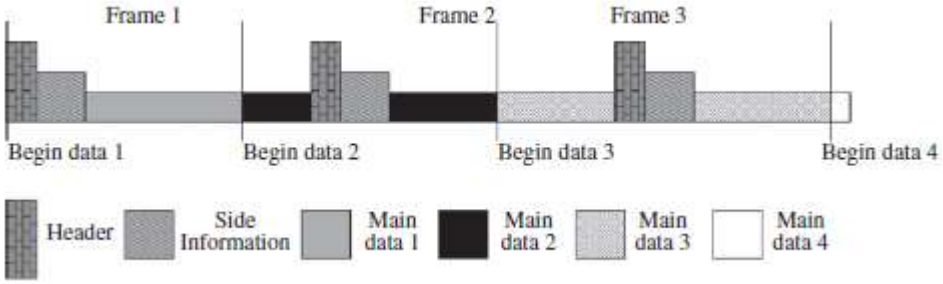
Последовательность окон



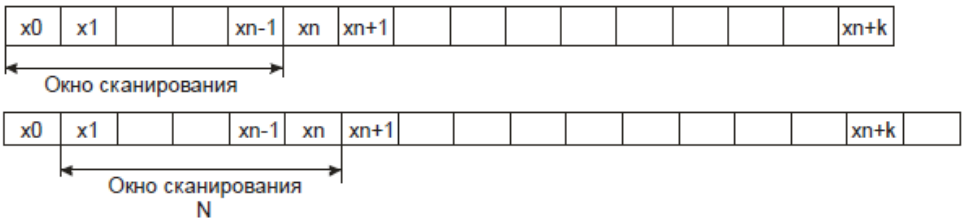




Frame structure for Layer 3. Sequence of windows



Отображение в частотную область



Кодер

1. Поместить 32 новых сэмпла в буфер FIFO X.

2. Окно сэмплов: $Z_i = C_i \times X_i$, для $i = 0, \ldots, 511$.

3. Промежуточное вычисление: $Y_i = \sum_{i=0}^{7} Z_{i+64j}$ для $i = 0, \ldots, 63$.

4. Вычисление 32 сигналов: $S_i = \sum_{k=0}^{63} M_{i,k} \times Y_k$ для $i = 0, \ldots, 31$.

5. Дать на выход 32 подполосных сигнала S_i.

$$S_t[i] = \sum_{k=0}^{63} \sum_{j=0}^{7} M_{i,k} \left(C[k+64j] \times X[k+64j] \right), \; i = 0, \ldots, 31.$$

$$M_{i,k} = \cos\left(\frac{(2i+1)(k-16)\pi}{64} \right), \; i = 0, \ldots, 31; \; k = 0, \ldots, 63.$$

Декодер

1. Прочитать 32 новых сигнала S_i, $i = 0, \ldots, 31$.

2. Сдвинуть* буфер FIFO $V_i = V_{i-64}$, $i = 1023, \ldots, 64$.

3. Умножить: $V_i = \sum_{k=0}^{32} N_{i,k} S_k$, $i = 0, \ldots, 63$.

4. Построить 512 компонентов вектора U:
Для $i = 0$ до 7 выполнить,
Для $j = 0$ до 31 выполнить,
$U_{j+64i} = U_{j+128i}$
$U_{32+j+64i} = U_{96+j+128i}$

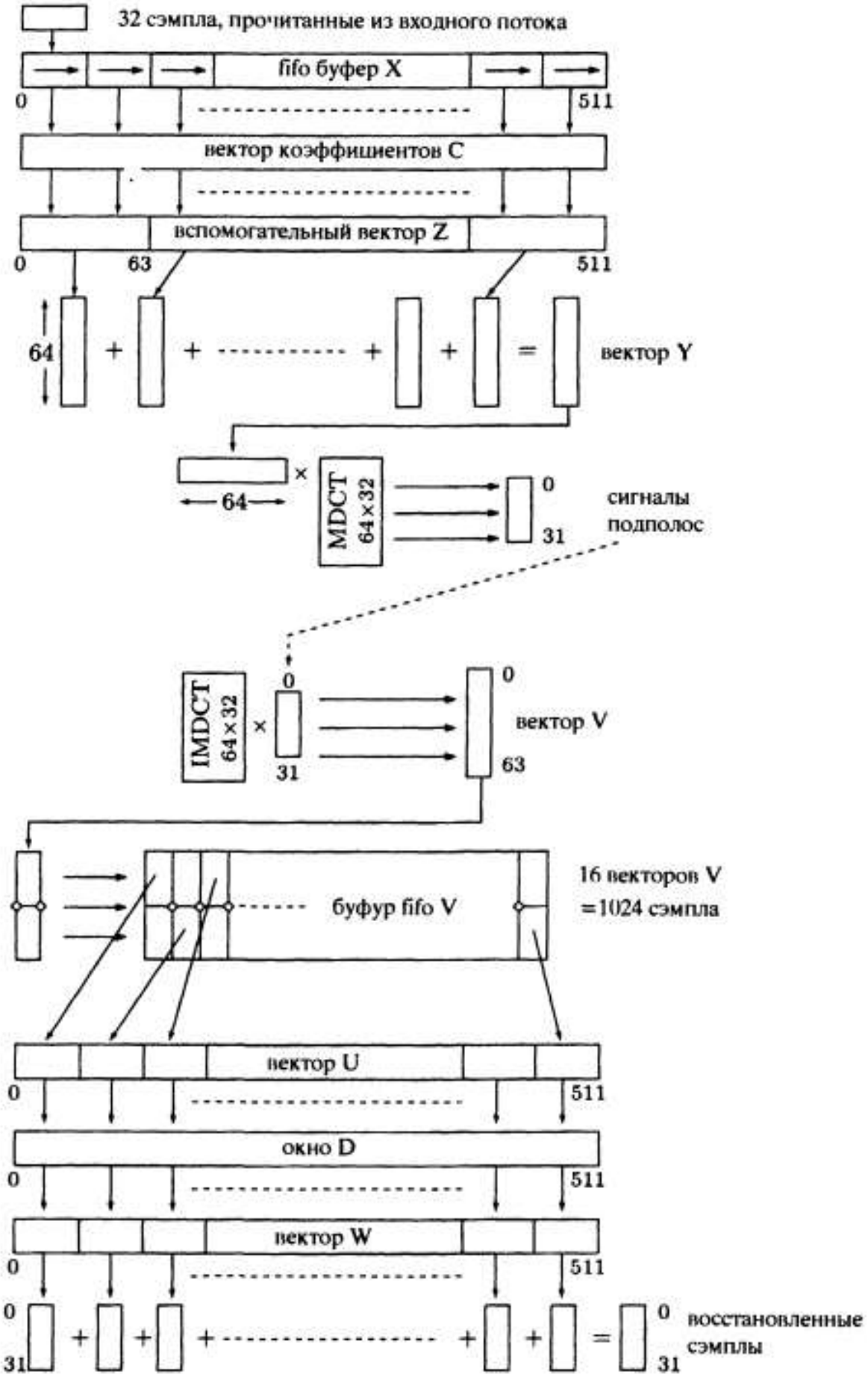
5. Построить W. Для $i = 0$ до 511 выполнить, $W_i = U_i \times D_i$.

6. Вычислить 32 сэмпла. Для $j = 0$ до 31 выполнить, $S_j = \sum_{i=0}^{15} W_{j+32i}$.

7. Выдать на выход 32 сэмпла S_j.

*В начале вектор V инициализируется нулями.

$$N_{i,k} = \cos\left(\frac{(2k+1)(i+16)\pi}{64} \right), \; i = 0, \ldots, 63; \; k = 0, \ldots, 31.$$


32 сэмпла, прочитанные из входного потока
fifo буфер X
0
511
вектор коэффициентов C
вспомогательный вектор Z
0
63
511
64
вектор Y
64
MDCT 64×32
0
31
сигналы подполос
IMDCT 64×32
0
31
0
63
вектор V
буфур fifo V
16 векторов V =1024 сэмпла
вектор U
0
511
окно D
0
511
вектор W
0
511
0
31
0
31
восстановленные сэмплы

Психоакустическая модель

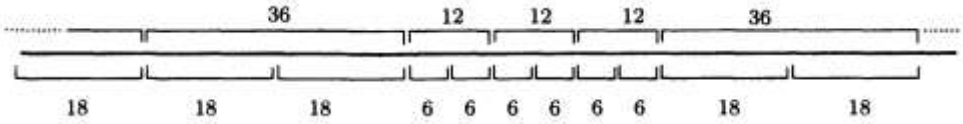
Основные шаги модели состоят в следующем:

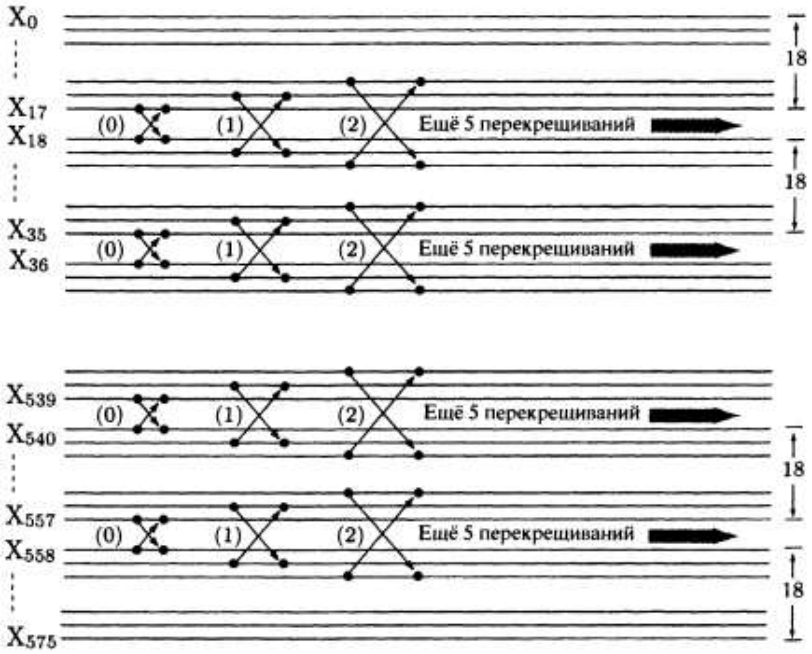
1. Спектральные значения частотных полос разделяются на тональные (подобные синусоиде) и нетональные (шумоподобные) компоненты.
2. Прореживается спектр тональных компонент. Исключаются компоненты, лежащие ниже абсолютного порога слышимости. Оставшиеся компоненты прореживаются с помощью "окна" шириной 0,5 Барк.
3. В каждом поддиапазоне вычисляется порог маскирования для тональных и не тональных компонент. Маскируемые компоненты отбрасываются.
4. Вычисляется общее отношение сигнал/маскирование (SMR) для каждой субполосы кодирования.

MDCT

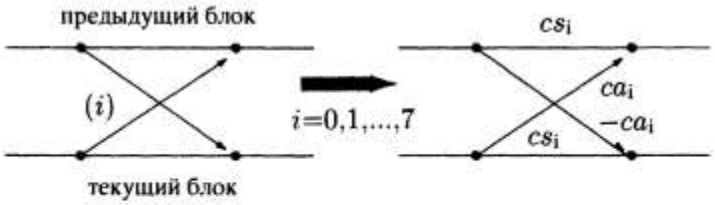
$$S_i = \sum_{k=0}^{n-1} x_k \cos\left(\frac{\pi}{2n}\left[2k+1+\frac{n}{2}\right](2i+1)\right),\ i=0,1,\ldots,\frac{n}{2}-1,$$

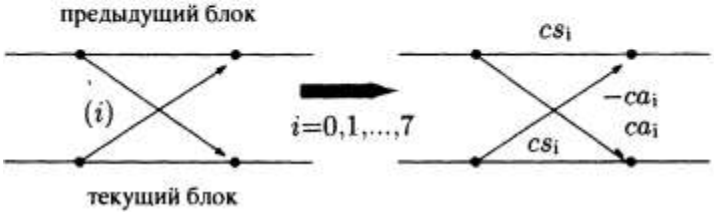
$$x_k = \sum_{i=0}^{n/2-1} S_i \cos\left(\frac{\pi}{2n}\left[2k+1+\frac{n}{2}\right](2i+1)\right),\ k=0,1,\ldots,n-1$$






```
for(sb=1; sb<32; sb++)
 for(i=0; i<8; i++){
  xar[18*sb-1-i]=xr[18*sb-1-i]cs[i]-xr[18*sb+i]ca[i]
  xar[18*sb+i]=xr[18*sb+i]cs[i]+xr[18*sb-1-i]ca[i]
                    }
```



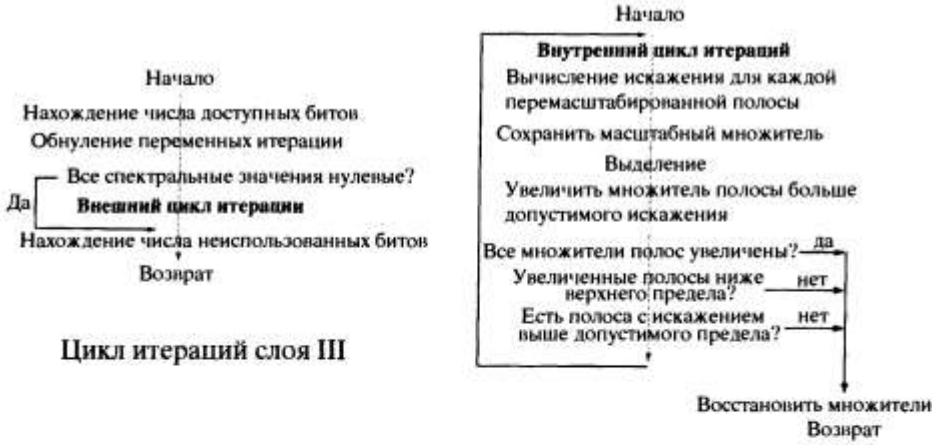


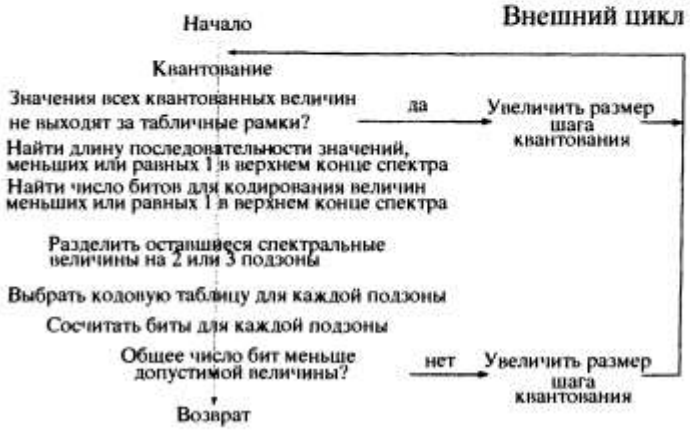
$$cs_i = \frac{1}{\sqrt{1+c_i^2}},\ ca_i = \frac{c_i}{\sqrt{1+c_i^2}},\ i = 0, 1, \ldots, 7.$$

$$c_i\ :\ -0.6,\ \ -0.535,\ \ -0.33,\ -0.185,\ \ -0.095,\ \ -0.041,\ \ \ -0.0142,\ \ -0.0037.$$

$$is(i) = \left[\left(\frac{xr(i)}{\text{quant}}\right)^{3/4} - 0.0946\right]$$

Цикл итераций слоя III





Внутренний цикл итераций слоя III

Другие аудиокодеры

AAC (Advanced Audio Coding)
MPEG II AAC
MPEG-4 AAC
Dolby AC-3