第三讲 音频压缩编码

- 音频压缩编码基本原理
- MPEG-1 音频压缩算法及标准
- MPEG-2 Audio
- MPEG-4 Audio
- AC-3音频编码

一、音频压缩编码基本原理

- **1**、什么是音频信号？
- 通常将人耳可以听到的频率在**20Hz**到**20KHz**的声波称为声音信号,声音振动被拾音器转换成电信号称为音频信号。
- 人的发音器官发出的声音频段在**80Hz**到**3400Hz**之间；
- 人说话的信号频率在**300Hz**到**3000Hz**，将该频段的信号称为语音信号。

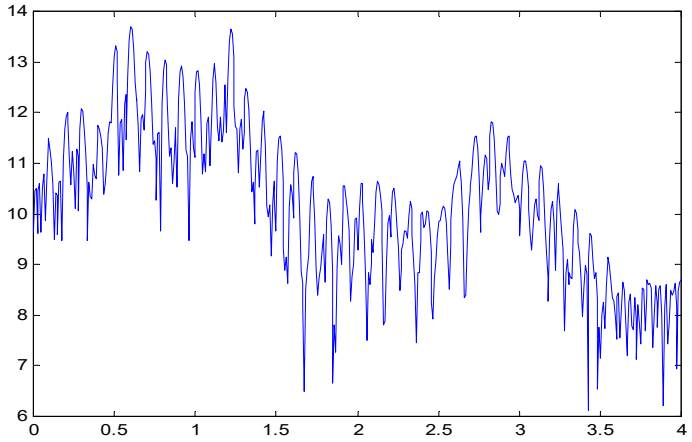
一、音频压缩编码基本原理

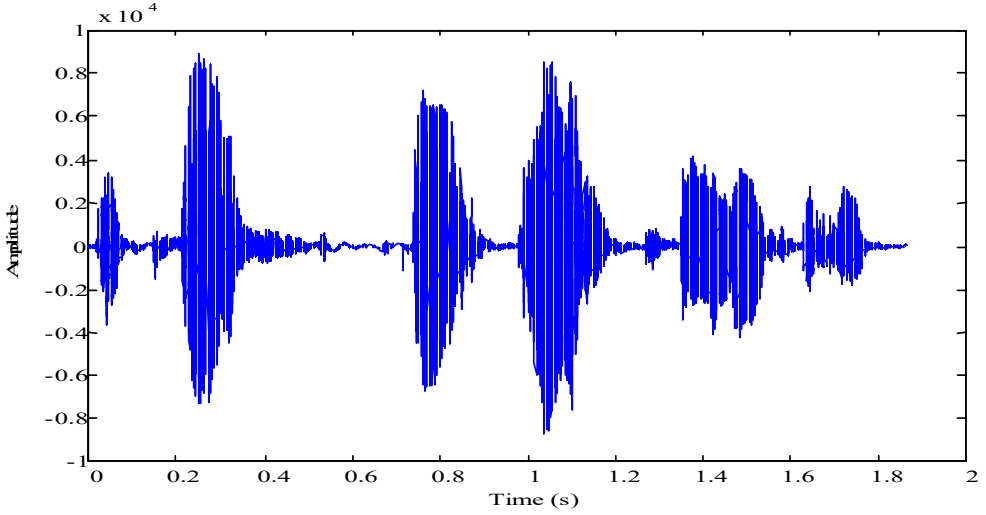
2、音频压缩的可能性

（1）声音信号中的“冗余”

频域：非均匀功率密度谱，低频能量高，高频能量低。

时域：信息冗余度主要表现在幅度非均匀分布，即不同幅度的样值出现的概率不同，小幅度的样值比大幅度样值出现的概率高。





一、音频压缩编码基本原理

2、音频压缩的可能性

（2）人耳的听觉特性，声音中存在与听觉无关的“不相关”部分。

对于人耳感觉不到的不相关部分不编码、不传送，以达到数据压缩的目的。

——利用了人耳听觉的心理声学特性。

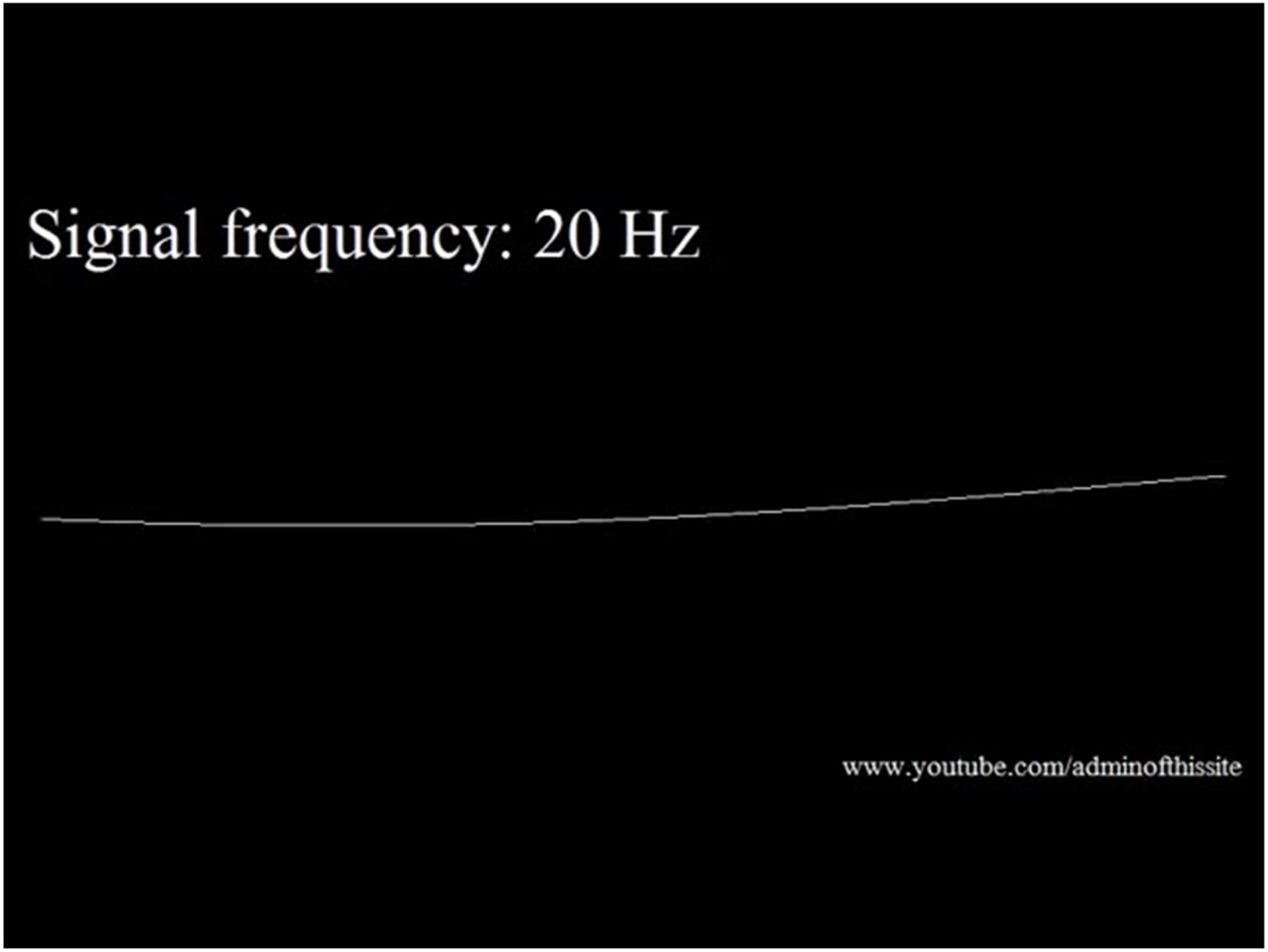
声音主观感受——响度、音调、音色；

声音客观特性——振幅、频率、频谱特性；

音色取决于不同的泛音，即声音的谐波部分。

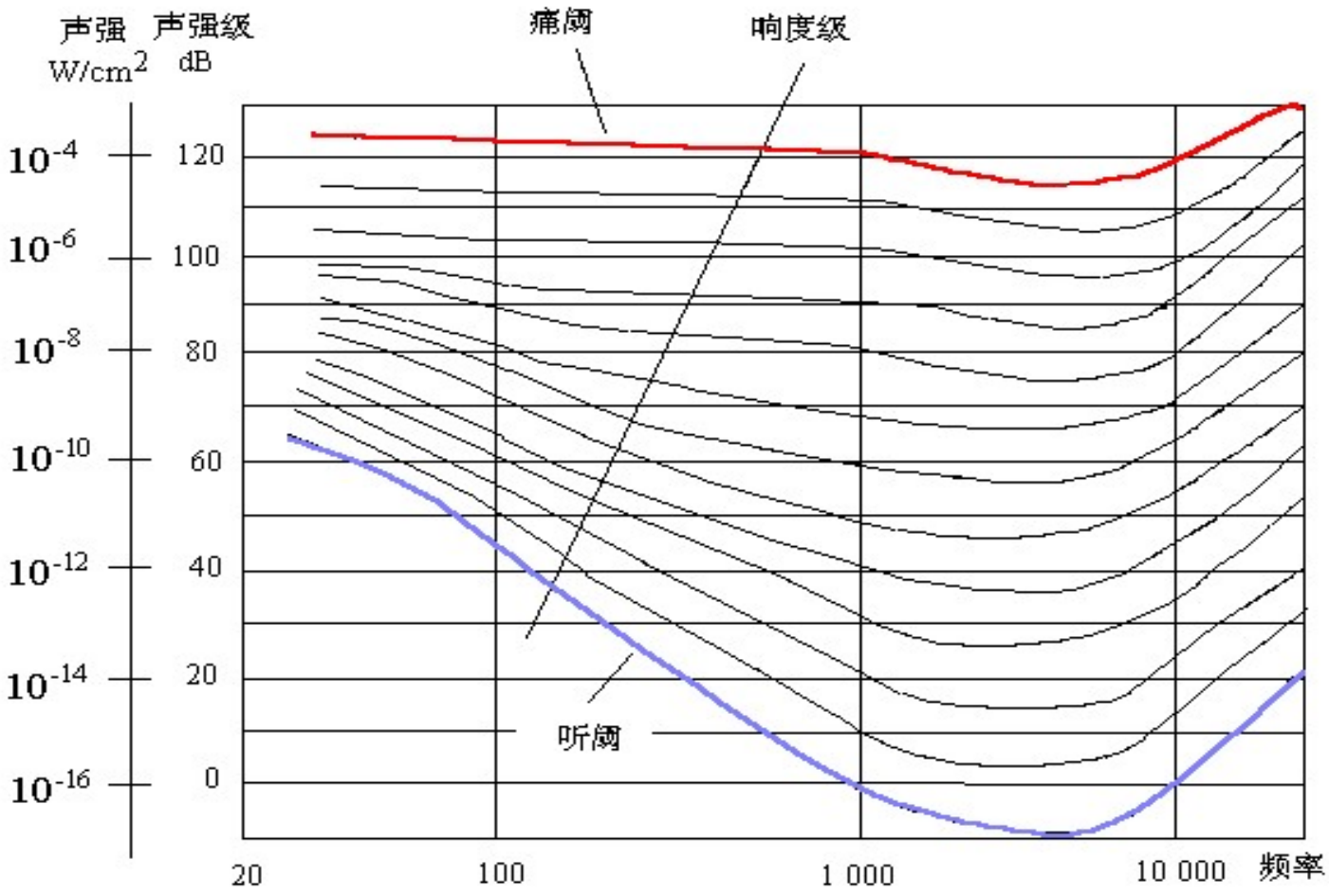
示例视频——**https://www.youtube.com/watch?v=qNf9nzvnd1k**

二、人类听觉系统的感知特性



二、人类听觉系统的感知特性

听阈 - 频率曲线



两个声音响度级相同，但强度不一定相同，还与频率有关；

声压级越高，等响度曲线趋于平坦；

人耳对**3~4KHz**的声音感觉最灵敏；

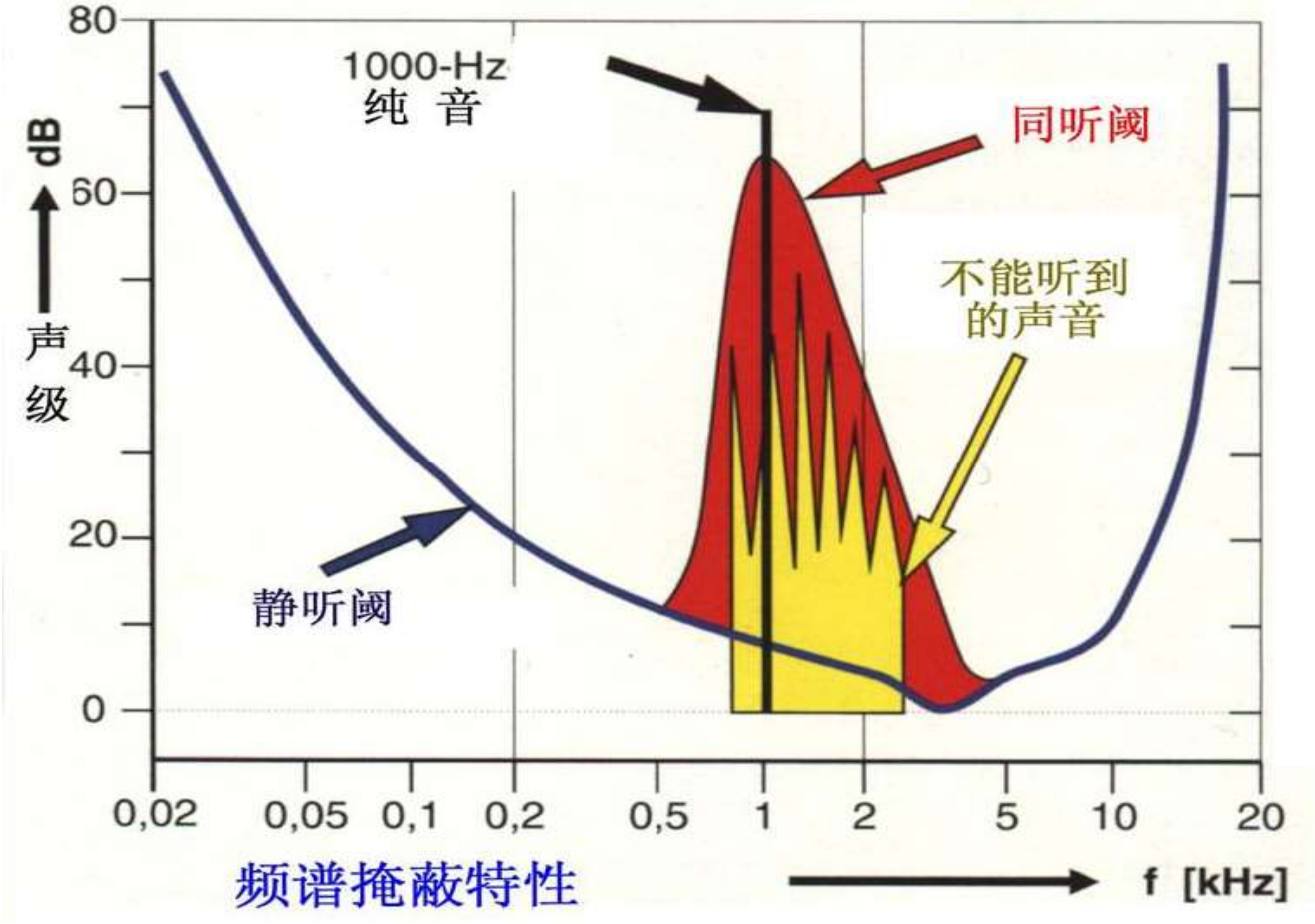
人耳的掩蔽效应

- 一个较弱的声音的听觉感受被另一个较强的声音影响的现象称为人耳的听觉掩蔽效应。听不到叫被掩蔽声，起掩蔽作用的叫掩蔽声。
- 被掩蔽音单独存在时的听阈分贝值，为绝对听阈。即安静环境中能被人耳听到的纯音最小值。也称静听域。
- 频域掩蔽/时域掩蔽。

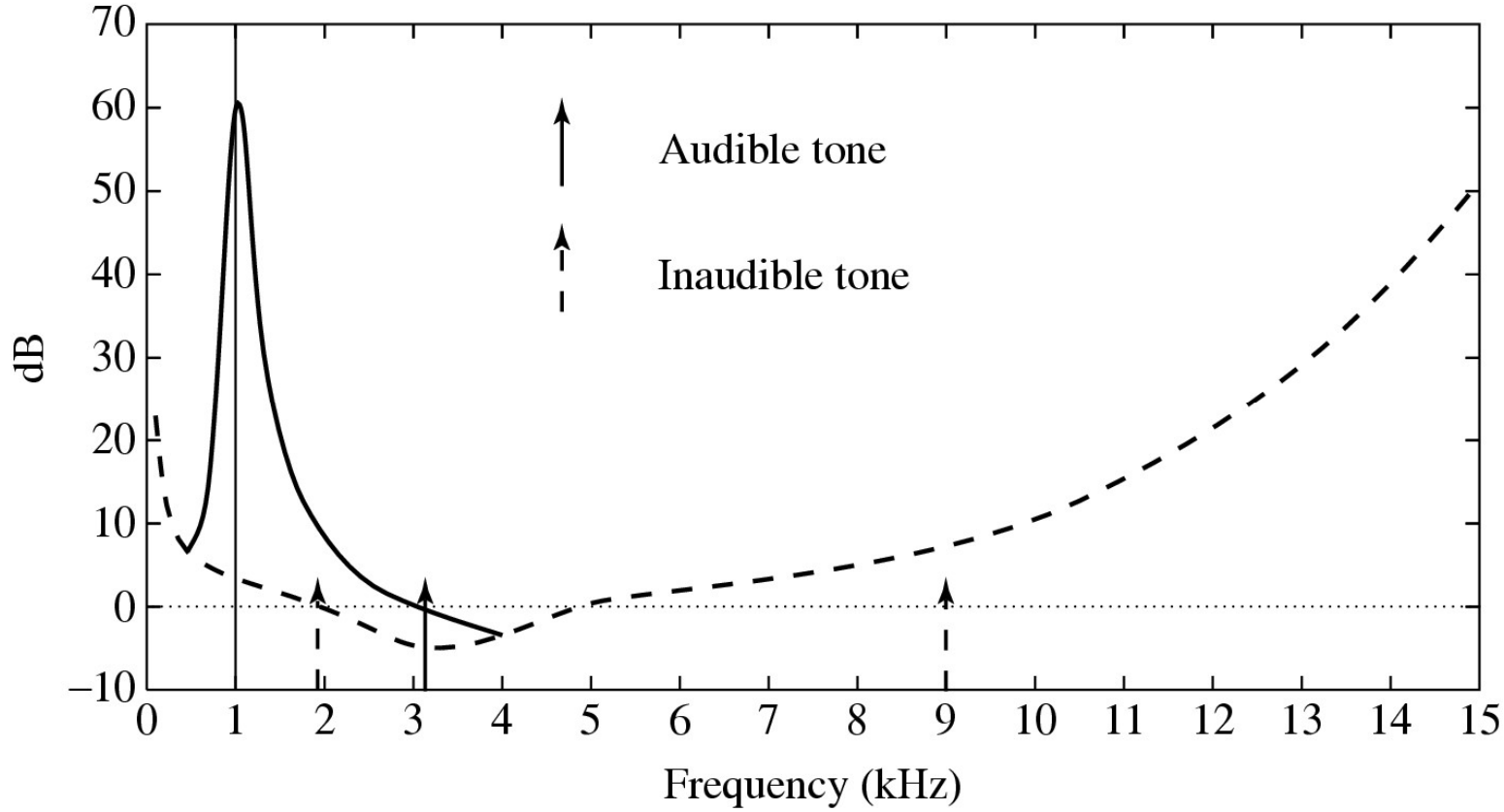
掩蔽效应演示——Simultaneous masking.mp4

1、频域掩蔽（纯音间的掩蔽）

一个强纯音会掩蔽在其附近同时发声的弱纯音，这种特性称为频域掩蔽，也称同时掩蔽。

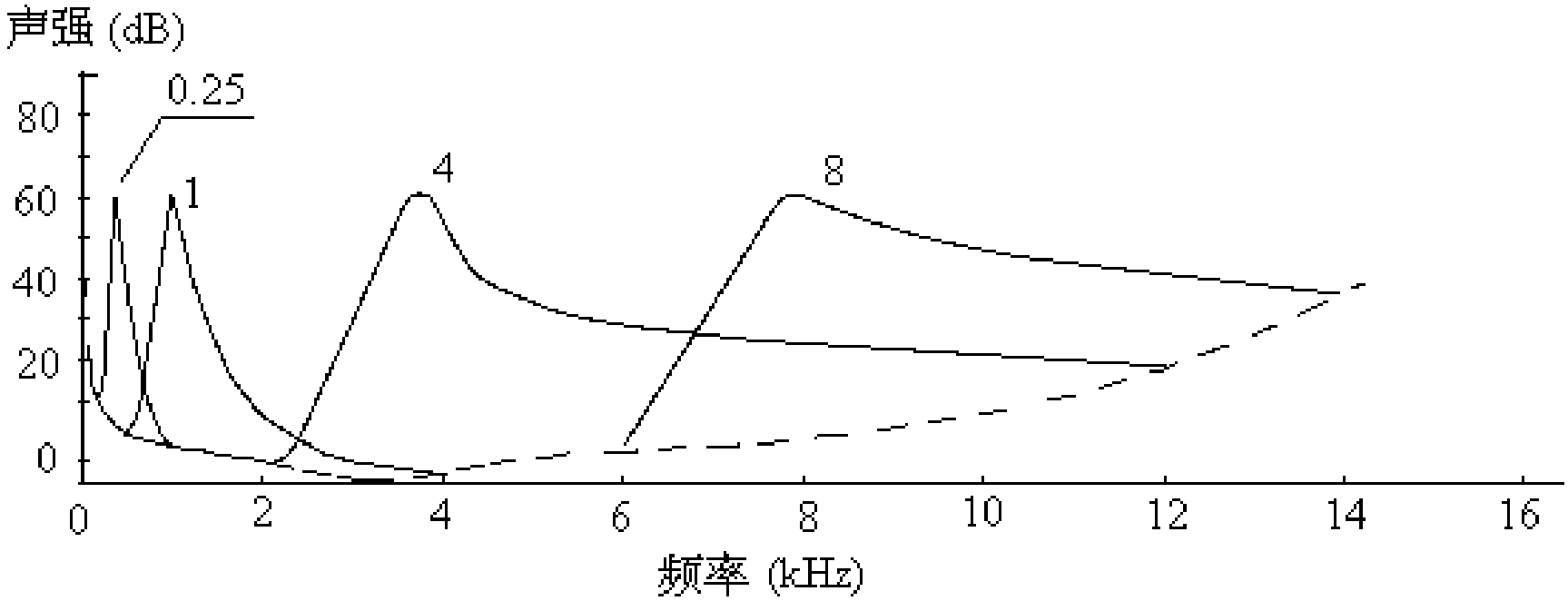


https://www.youtube.com/watch?v=2HDka1hYiCk



Effect on threshold for 1 kHz masking tone

频域掩蔽域随频率变化曲线

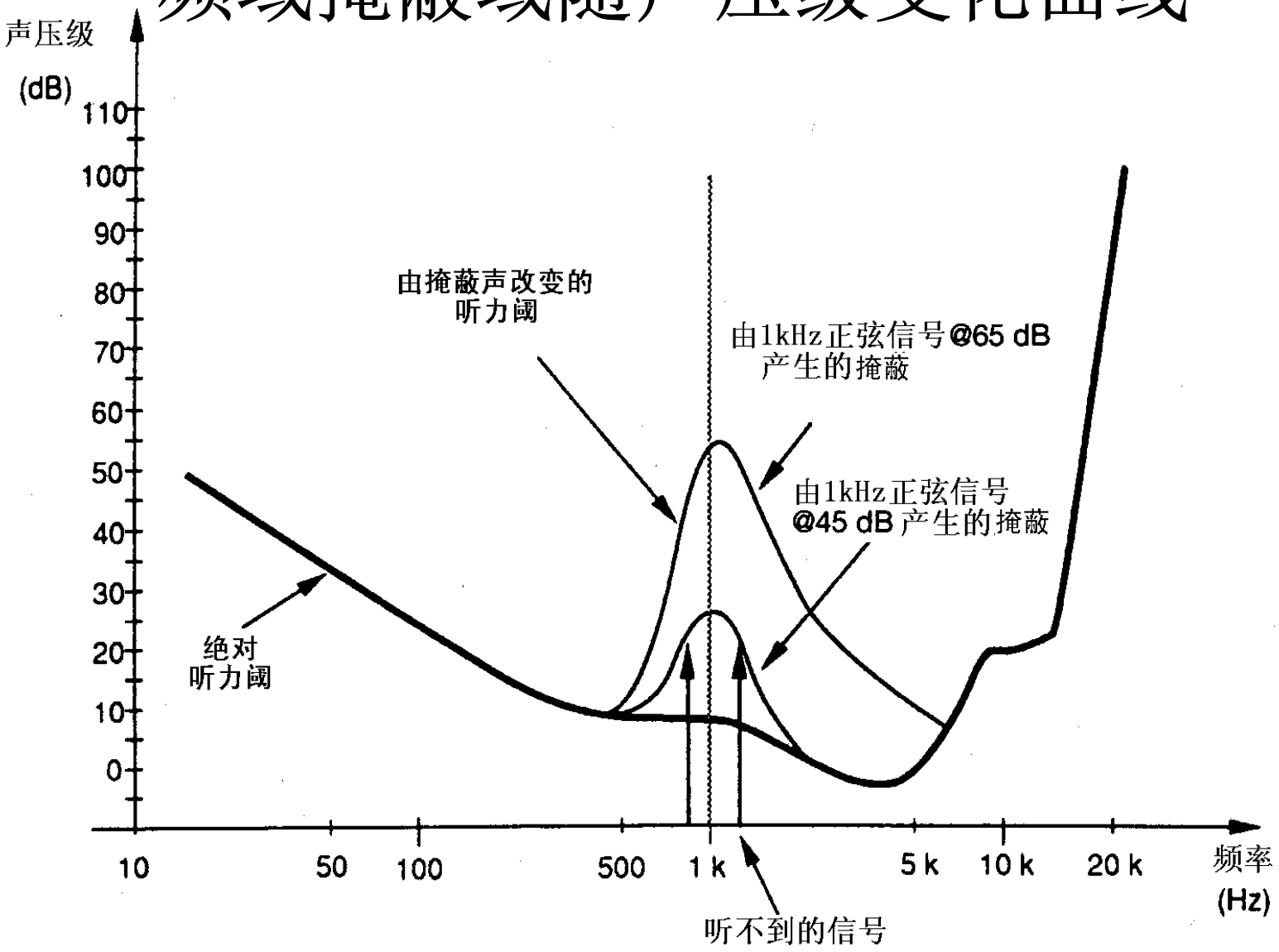


音调音的掩蔽阈的宽度随频率而变化；

掩蔽曲线不对称，高频段一侧的曲线斜率缓些；

低频音容易对高频音产生掩蔽。

频域掩蔽域随声压级变化曲线

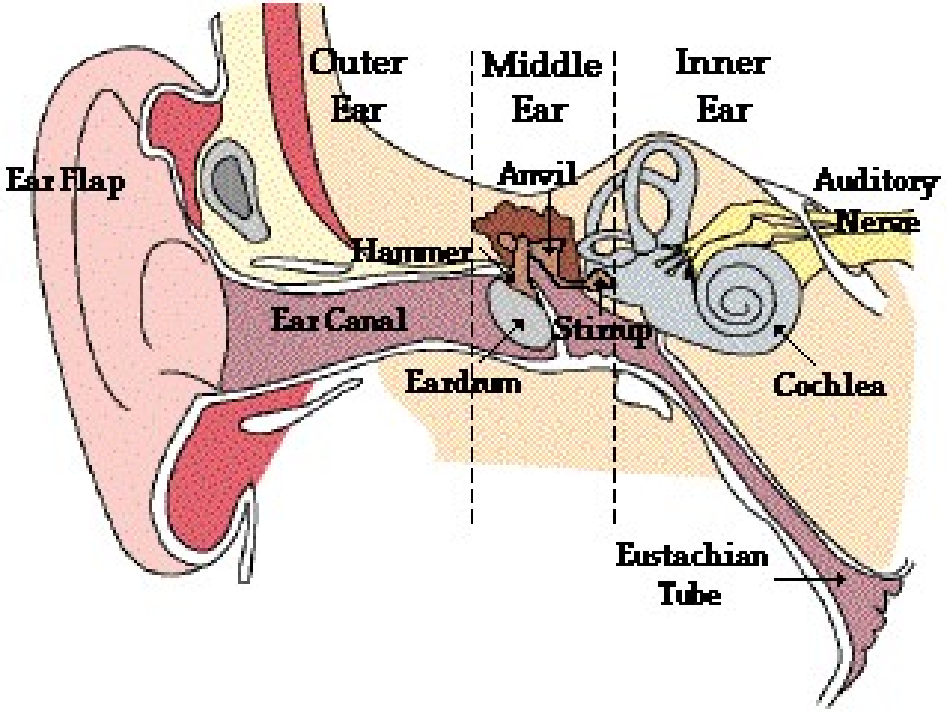


2、人耳模型——How ear works视频演示

2、人耳模型——Cochelar animation演示

2、人耳模型

- 声音频率发生转换
  - 声波冲击耳鼓（Eardrum)和连着的耳骨；
  - 耳鼓和耳骨将机械振动传递给耳蜗（Cochlea）
  - 耳蜗薄膜的椭圆窗沿基底膜长度方向引导行波；
  - 行波在薄膜的特定频率感应位置产生峰值响应；
  - 薄膜的特定频率感应位置为特定频带提供峰值响应；
- 可以把耳蜗当成一组高度重叠的带通滤波器



人耳相当于一个滤波器组

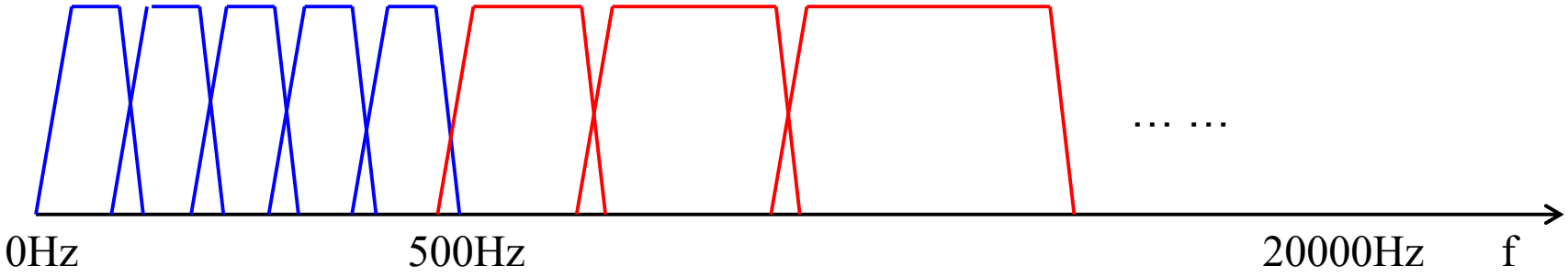
- 人类听觉系统大致等效于一个在0Hz到20KHz频率范围内由25个重叠的带通滤波器组成的滤波器组。

– 人耳不能区分同一频带内同时发生的不同声音；

– 人耳频带被称为临界频带（critical band）；

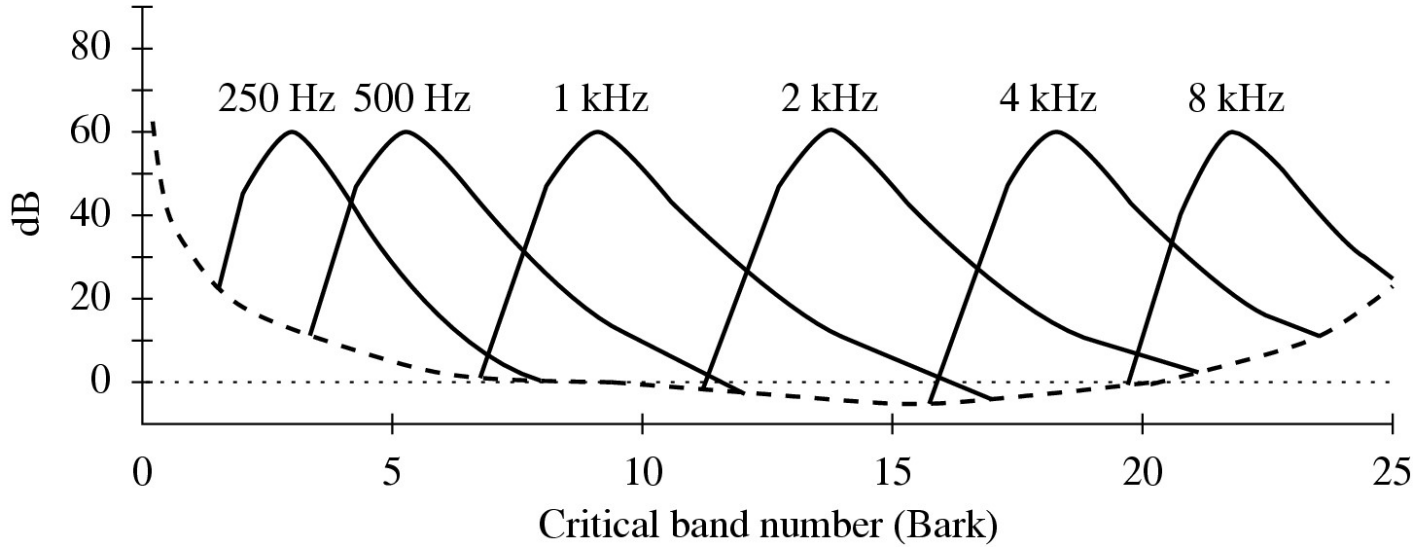
– 500Hz以下每个临界频带的带宽大约是100Hz，从500Hz起，临界频带带宽线性增加。

– 一个临界频带的带宽单位为1巴克（bark)。



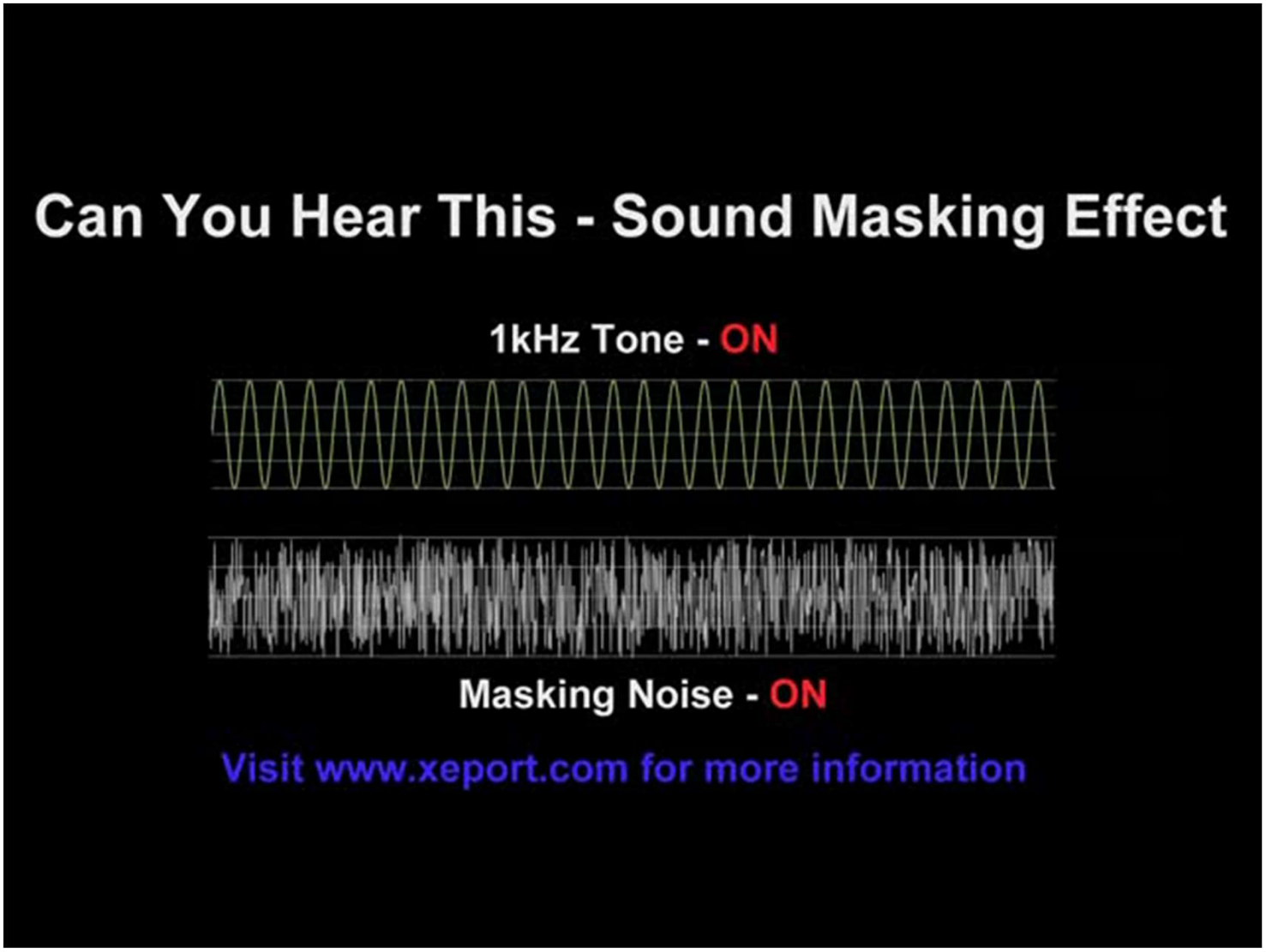
临界频带单位巴克（Bark）

- 对于任何掩蔽频率，巴克被定义为一个临界频带的宽度；
- 巴克单位的意义: 用巴克来衡量每个临界频带的宽度大致都是相同的。



用巴克单位表示的声音掩蔽效应

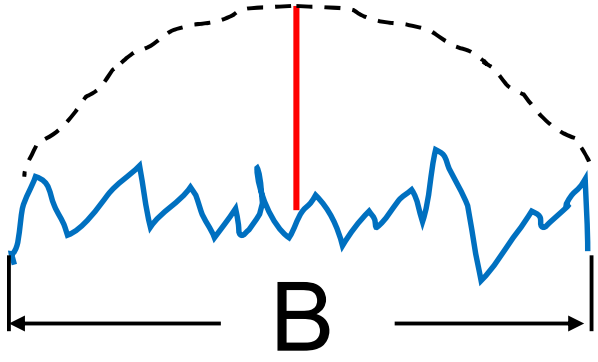
3、临界频带——噪声对纯音的掩蔽



3、临界频带——噪声对纯音的掩蔽

- 临界频带是指当某个纯音被以它为中心频率、且具有一定带宽的连续噪声所掩蔽时，如果该纯音刚好被听到时的功率等于这一频带内的噪声功率，这个带宽为临界频带宽度。
- 掩蔽效应在一定频率范围内不随带宽增大而改变，直至超过某个频率值。
- 通常认为从**20Hz**到**16kHz**有**25**个临界频带，单位为**bark**。
- **1bark** =一个临界频带的宽度
- **f< 500Hz** 时 **1bark**约为**f/100;**
- **f >500Hz** 时

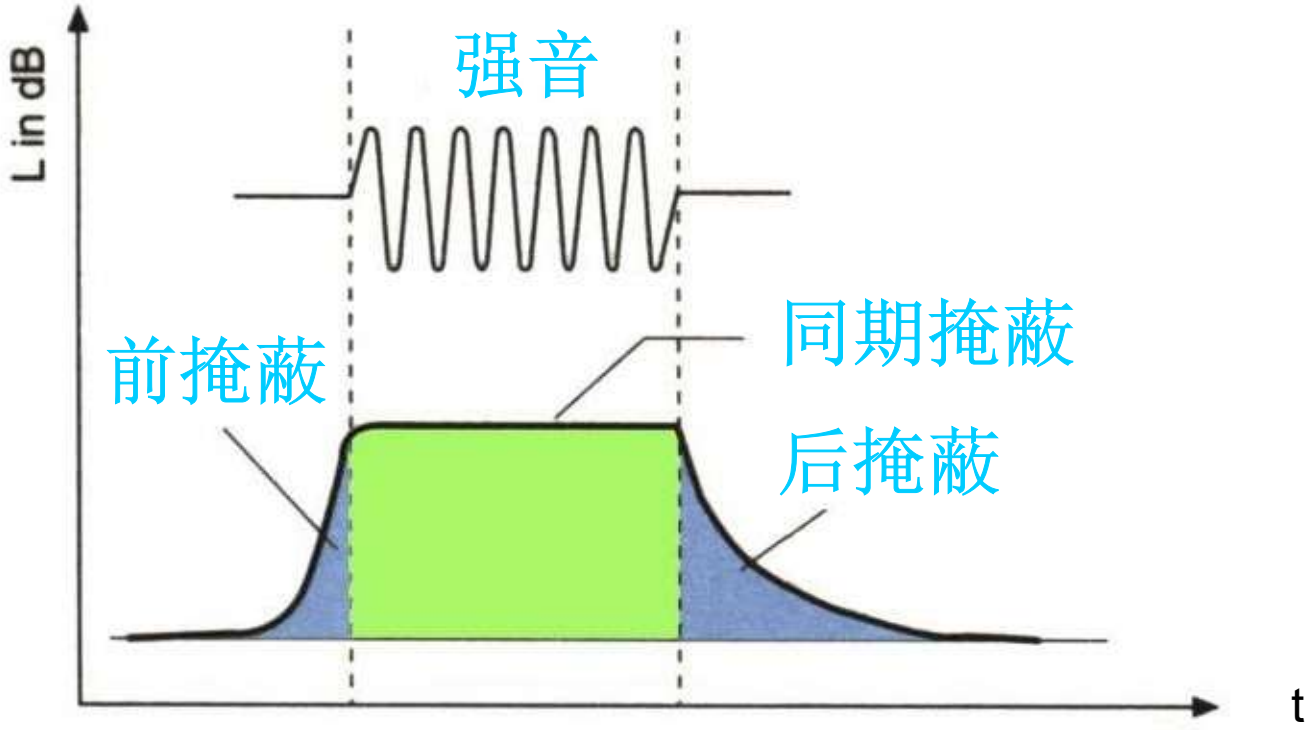
$$B = 13\,tan^{-1}\left(\frac{0.76f}{1000}\right) + 3.5\,tan^{-1}\left(\frac{f}{7500}\right)^2$$



临界	频率 (Hz)			临界	频率 (Hz)		
频带	低端	高端	宽度	频带	低端	高端	宽度
0	0	100	100	13	2000	2320	320
1	100	200	100	14	2320	2700	380
2	200	300	100	15	2700	3150	450
3	300	400	100	16	3150	3700	550
4	400	510	110	17	3700	4400	700
5	510	630	120	18	4400	5300	900
6	630	770	140	19	5300	6400	1100
7	770	920	150	20	6400	7700	1300
8	920	1080	160	21	7700	9500	1800
9	1080	1270	190	22	9500	12000	2500
10	1270	1480	210	23	12000	15500	3500
11	1480	1720	240	24	15500	22050	6550
12	1720	2000	280				

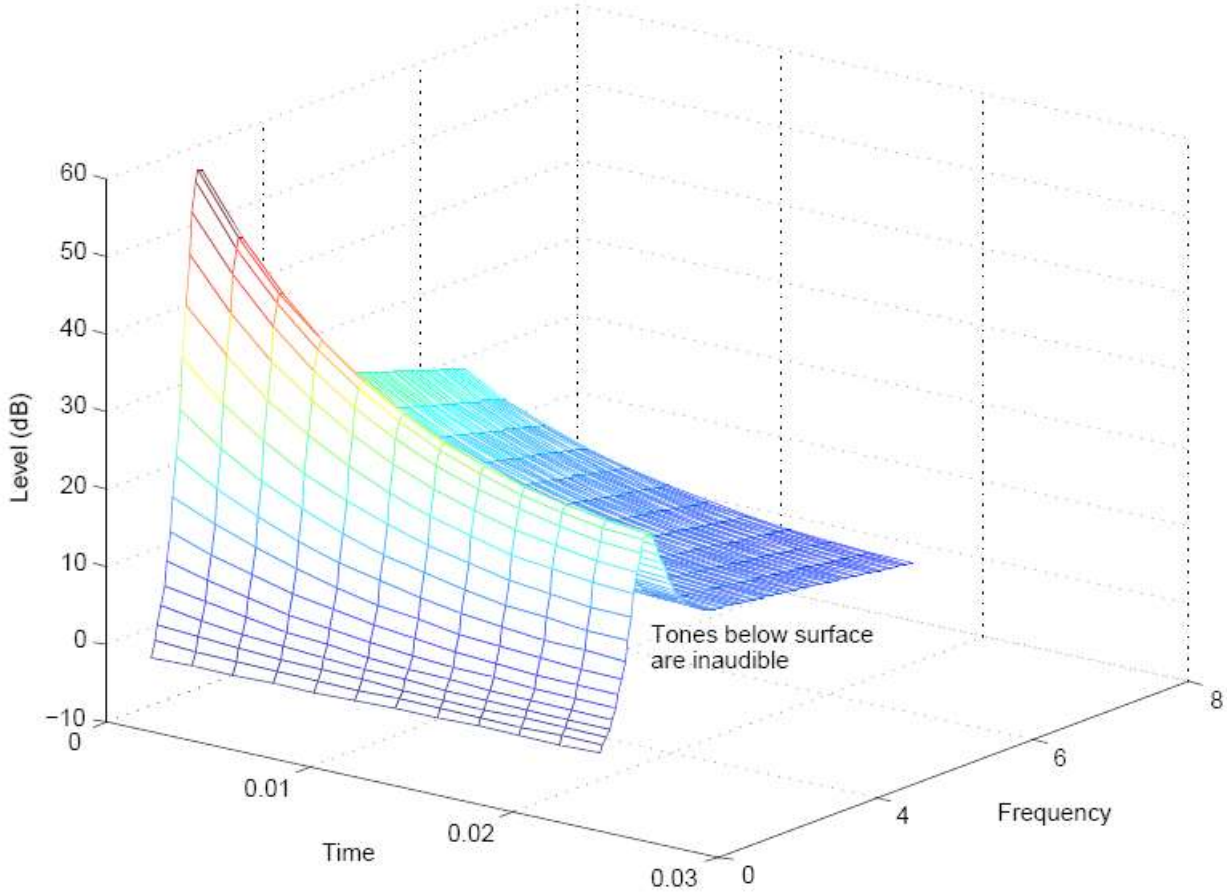
4、时域掩蔽

在时间上相邻的声音之间也有掩蔽现象。时域掩蔽又分为超前掩蔽和滞后掩蔽。超前掩蔽很短，只有大约**5～20 ms**，而滞后掩蔽可以持续**50～200 ms**。



时间掩蔽利用

- 基于时间掩蔽效应的编码策略是，编码时将时间上相继的一些样值归并成块，并计算每块内最大样值的比例因子；
- 据心理声学的掩蔽模型，对同一子带内相邻三个比例因子，可丢弃较小的因子，以减少传输比例因子的比特数。



Effect of temporal and frequency masking depending on both time and closeness in frequency.

音频信号幅度与编码的关系

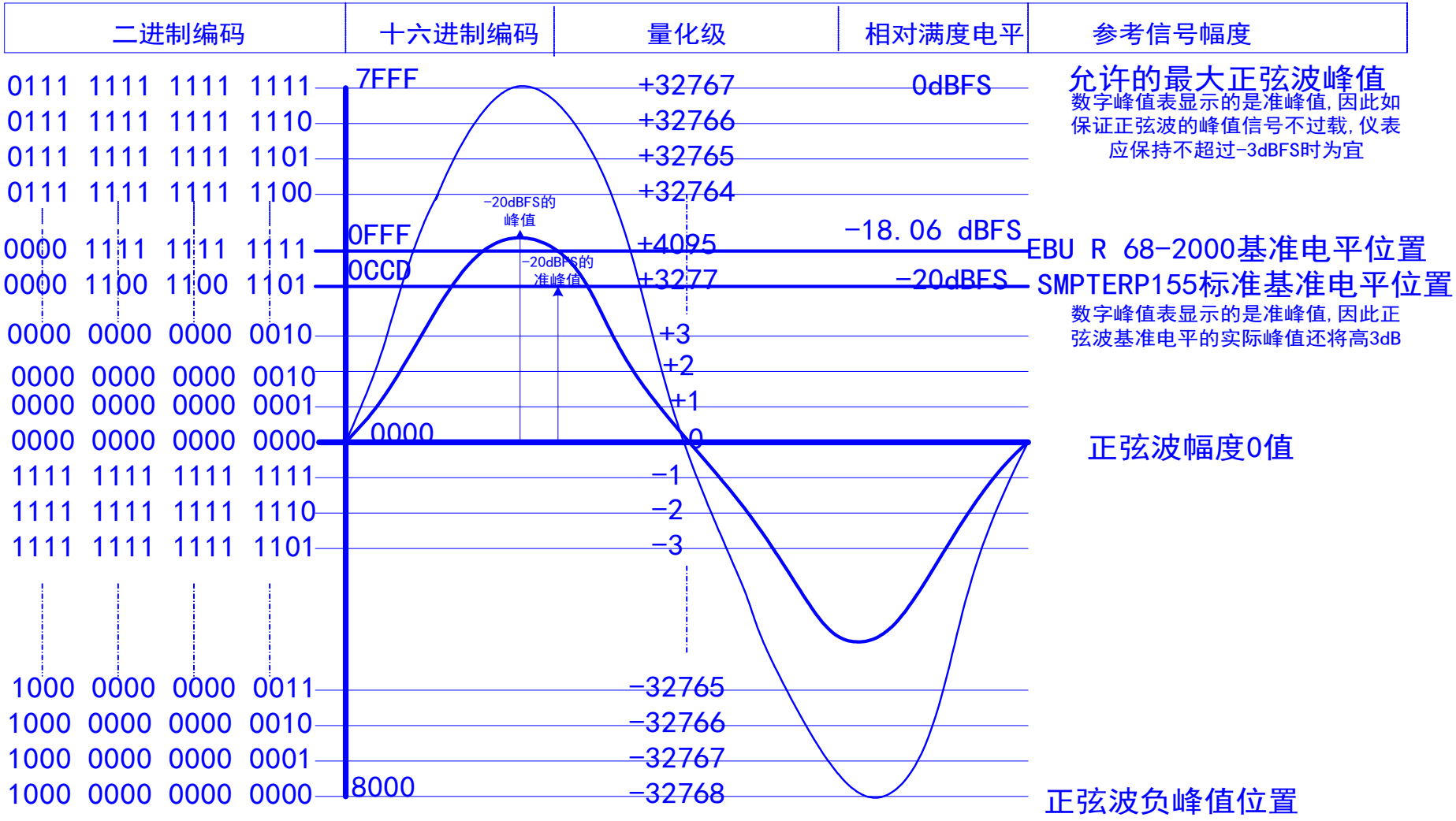


图6 16比特有效位编码的二进制、十六进制编码、量化级和相对满度电平的对应关系

得到音频信号幅度与编码的关系

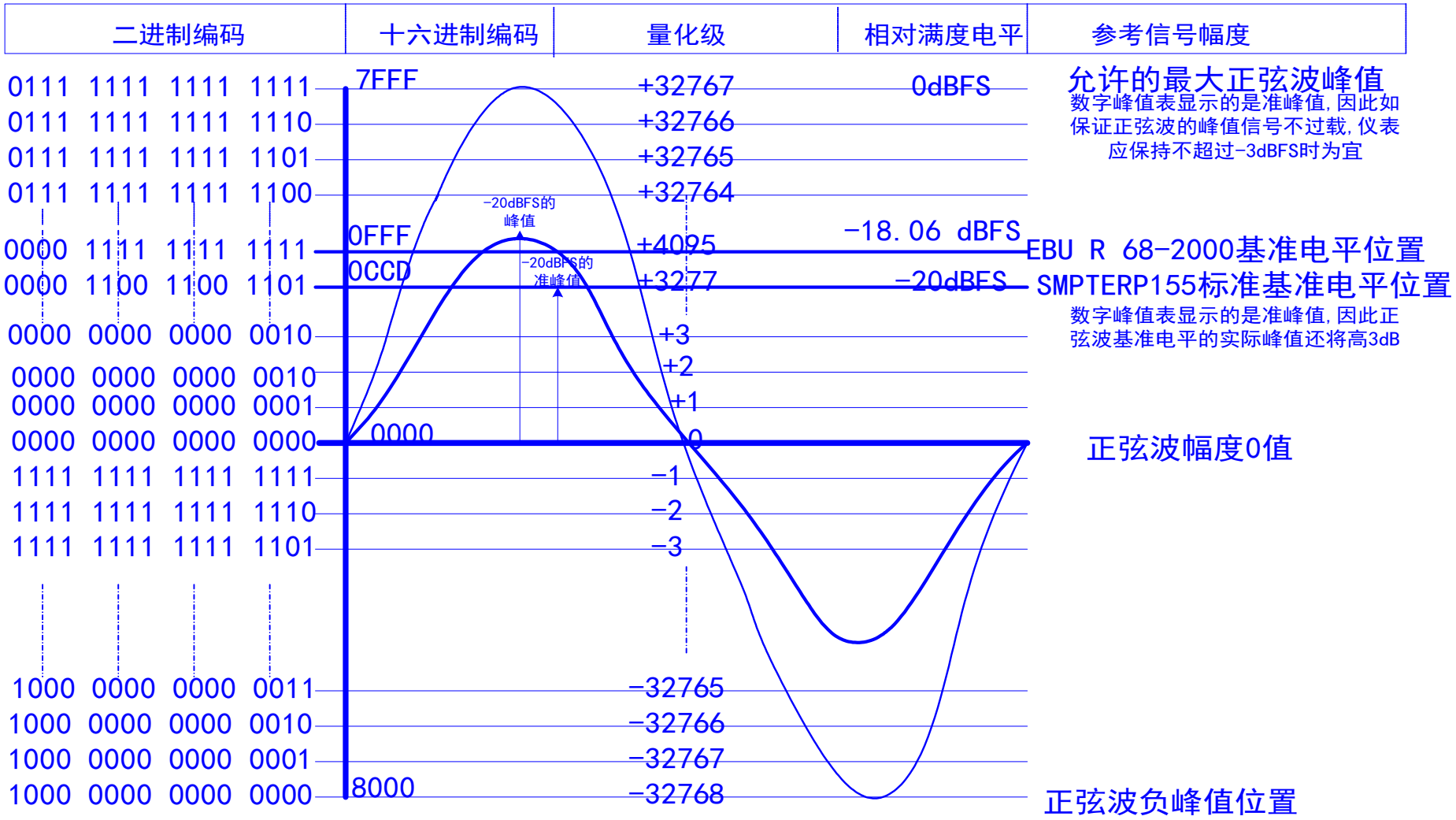


图6 16比特有效位编码的二进制、十六进制编码、量化级和相对满度电平的对应关系

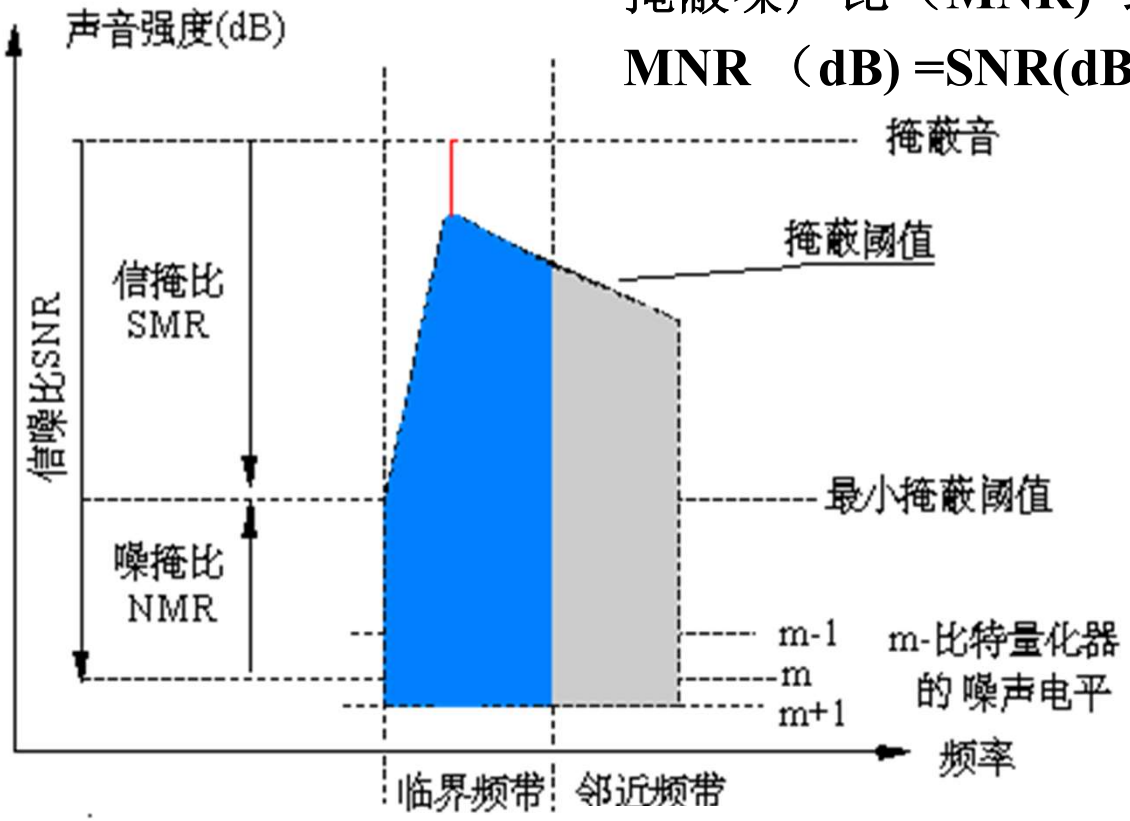
音频压缩处理相关结论

信噪比（SNR)=信号峰值—噪声有效值

信号掩蔽比（SMR)=信号峰值—最小掩蔽阈值

掩蔽噪声比（MNR)=最小掩蔽阈值 —量化噪声

MNR （dB) =SNR(dB)— SMR(dB)

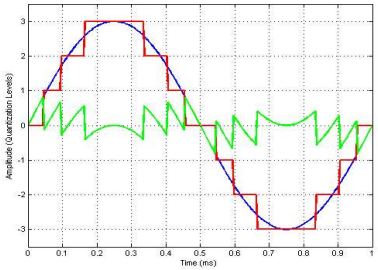


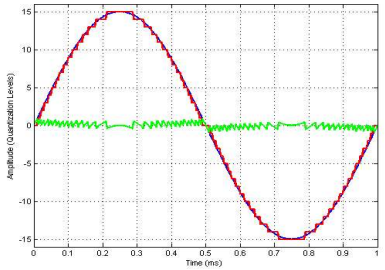
信噪比(SNR)=20lgL/N

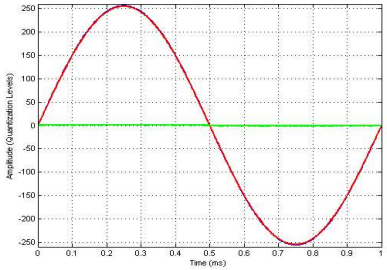
信噪比(SNR)=6.02n+1.76

N:量化噪声电平,n:量化比特数

重要结论：量化比特数增加1，量化信噪比提高6dB。

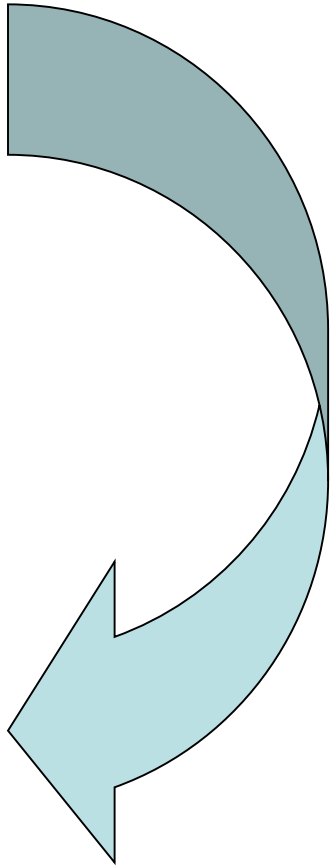






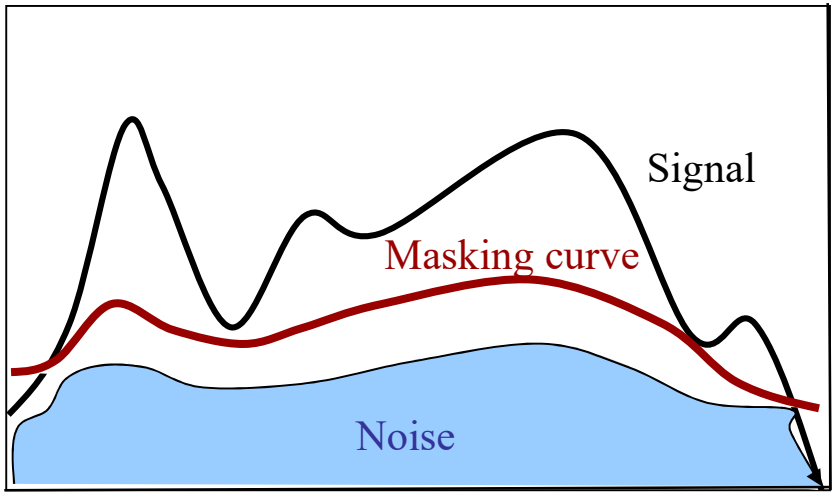
5、感知编码器原理

- 放弃物理上的同一性
- 得到感知上的同一性



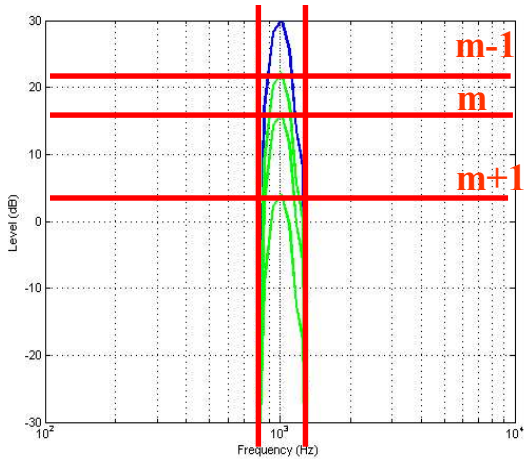
降低数据率

掩蔽的用途



- 去除会被掩蔽的信号分量
  - 因为即使传输了也不会被听见
    - 同听阈以下的信号部分不能被人耳听到（称不相关部分），不必传送。（去除不相关部分）
- 不理会可能被掩蔽的量化噪声
  - 因为会被信号淹没
    - 按同听阈以上的信号值计算量化比特数，对信号重新量化，使量化噪声在同听阈以下即可。

Example



- 假设32个子带中的16格子带如下：

子带	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
电平	0	8	12	10	6	2	10	60	35	20	15	2	3	5	3	1	(dB)

- 如果第**8**子带的电平是**60dB**，对第**7**子带产生**12dB**掩蔽，对第**9**子带产生**15dB**掩蔽；
- 第**7**子带电平为**10dB(<12dB)**，不必编码；
- 第**9**子带电平为**35dB(>15dB)**，需要编码；

6、音频信号压缩编码方法

（1）波形编码 ——直接对时域或频域波形编码

PCM , DPCM, ADPCM ,

子带编码,自适应变换编码

（2） 参数编译码器

从语音波形信号中提取语音生成模型的参数，使用这些参数通过语音生成模型重构出语音。

（3）混合编码

（4） 感知编码

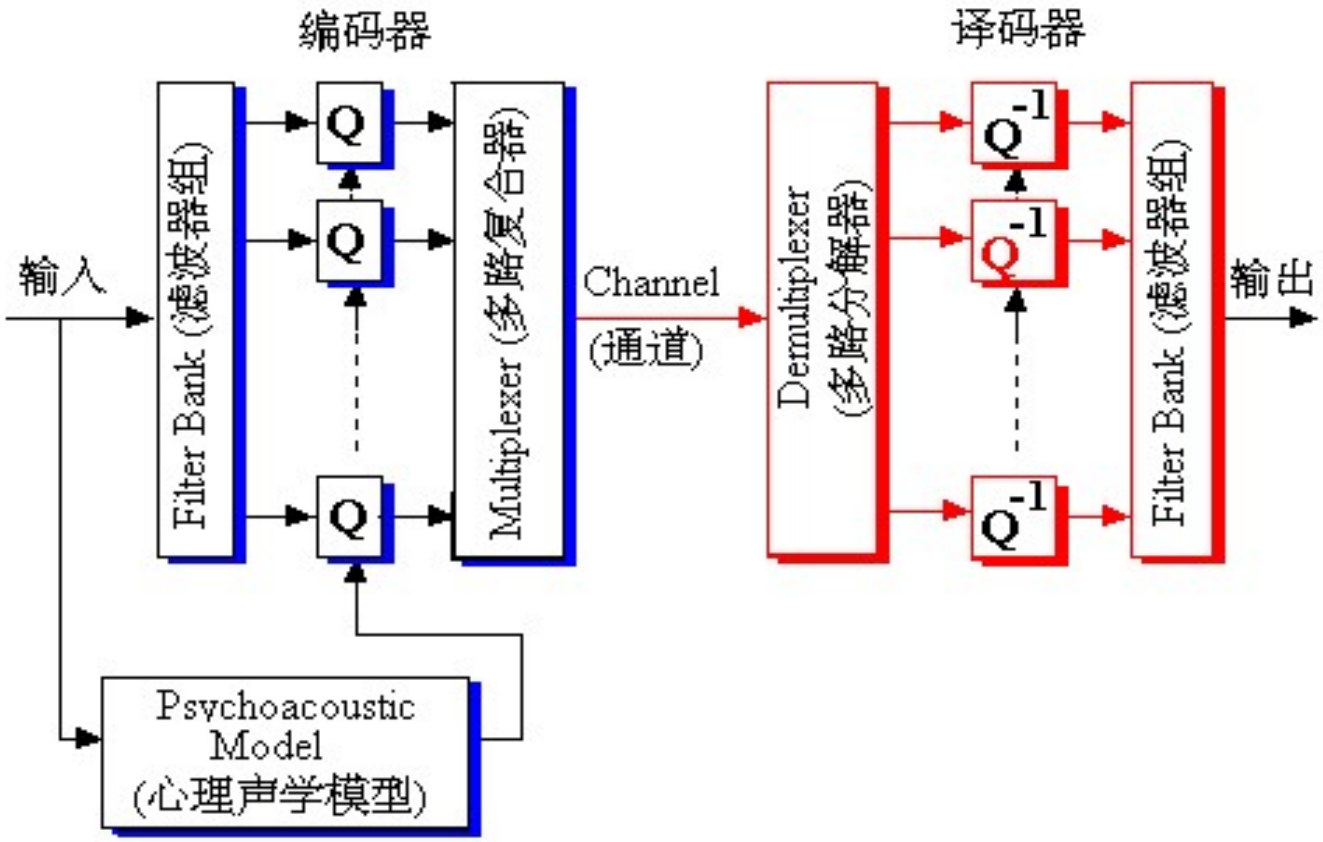
三、子带编码

子带编码(sub-band coding，SBC)

- 基本思想：使用一组带通滤波器(band-pass filter，BPF)把输入音频信号的频带分成若干个连续的频段，每个频段称为子带。对每个子带中的音频信号采用单独的编码方案去编码。在信道上传送时，将每个子带的代码复合起来。在接收端解码时，将每个子带的代码单独解码，然后把它们组合起来，还原出原来的音频信号。

1、感知子带压缩算法

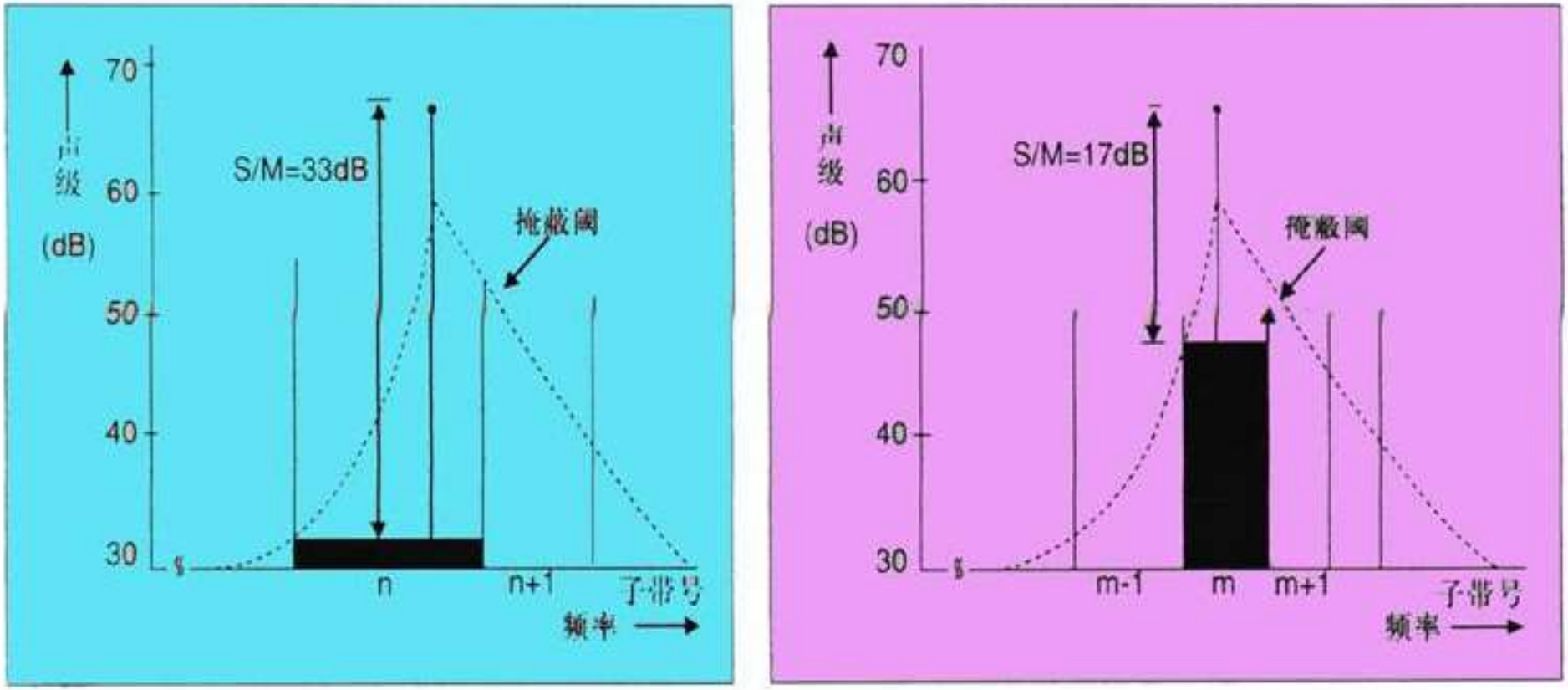
以心理声学模型为基础，主要利用了听觉阈值和听觉掩蔽特性。



1、感知子带压缩算法

- 用多相滤波器组，将宽带声音信号分割为多个子频带，对各子带的音频样值分别进行压缩编码。
- 理想的频带的分割应模仿临界频带，各子带的宽度不一致，随着频率的升高，子带的带宽也增加。
- 每个子带内根据信号掩蔽比确定样值的量化级数，量化噪声的高度与带内同听阈值越接近，数据率压缩越充分。
- 子带越多（越窄），在相同音质下编码所得数据率越低；传输中的比特差错仅限制在很窄的子频带内，影响越小。

窄子带能改善声音质量



不同子带宽度对比

2、子带编码的好处

第一，对每个子带信号分别进行自适应控制，量化阶的大小可以按照每个子带的能量电平加以调节。

第二，可根据每个子带信号在感觉上的重要性，对每个子带分配不同的位数，用来表示每个样本值。

例如，在低频子带中，为了保护音调和共振峰的结构，就要求用较小的量化阶、较多的量化级数，即分配较多的位数来表示样本值。而话音中的摩擦音和类似噪声的声音，通常出现在高频子带中，对它分配较少的位数。

3、MUSICAM编码

- **MUSICAM (Masking pattern adapted Universal Subband Integrated Coding And Multiplexing)**

  —掩蔽型自适应通用子带综合编码与复用。

  编码将宽带的音频信号频谱分为宽度为**750Hz**的**32**个子带，利用人耳听觉的心理声学现象和音频信号统计的内在联系，确定音频信号中的不相关部分和去除冗余，实现数据压缩。

- 一套**CD**立体声数据率为 **1411.2kbps**，**MUSICAM**编码后数据率为**2×96kbps**，重放仍有**CD**质量.

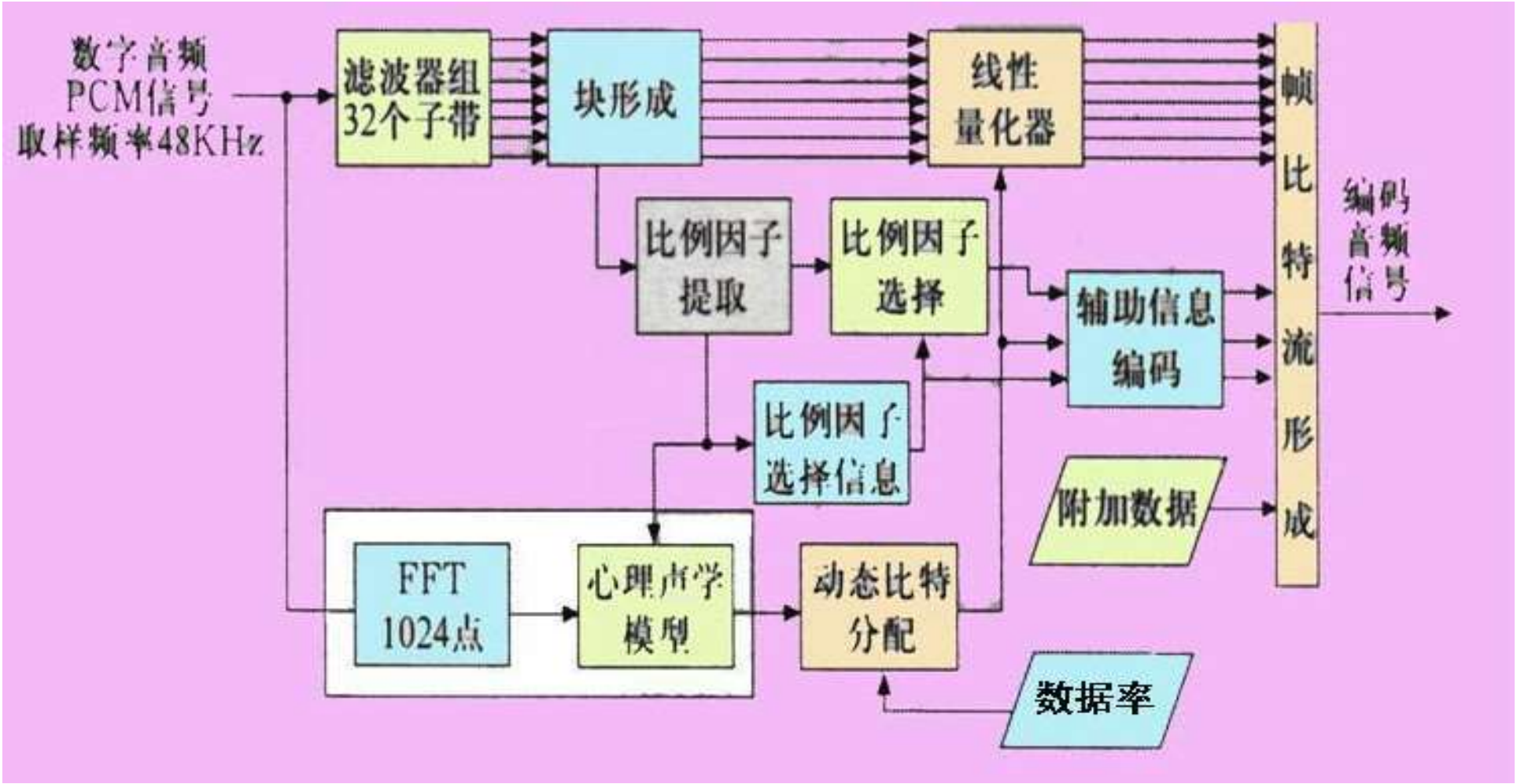
3、MUSICAM编码

- **MUSICAM与MPEG-1的Layer II一致；**

  Layer I 是MUSICAM的简化版本；

  Layer III 是MUSICAM与ASPEC（自适应谱感知熵编码）变换编码的结合，低比特率时质量最好，时域到频域的滤波器组提供了高频谱分辨率。在低码率(64 kbit/s)时，ASPEC表现出更为出色的音质，而MUSICAM则在编码解码的复杂度和延时上略胜一筹。

MUSICAM编码器


数字音频
PCM信号
取样频率48KHz
滤波器组
32个子带
块形成
线性
量化器
帧
比
特
流
形
成
编码
音频
信号
比例因子
提取
比例因子
选择
辅助信息
编码
比例因子
选择信息
附加数据
FFT
1024点
心理声学
模型
动态比特
分配
数据率

四、 音频压缩的国际标准

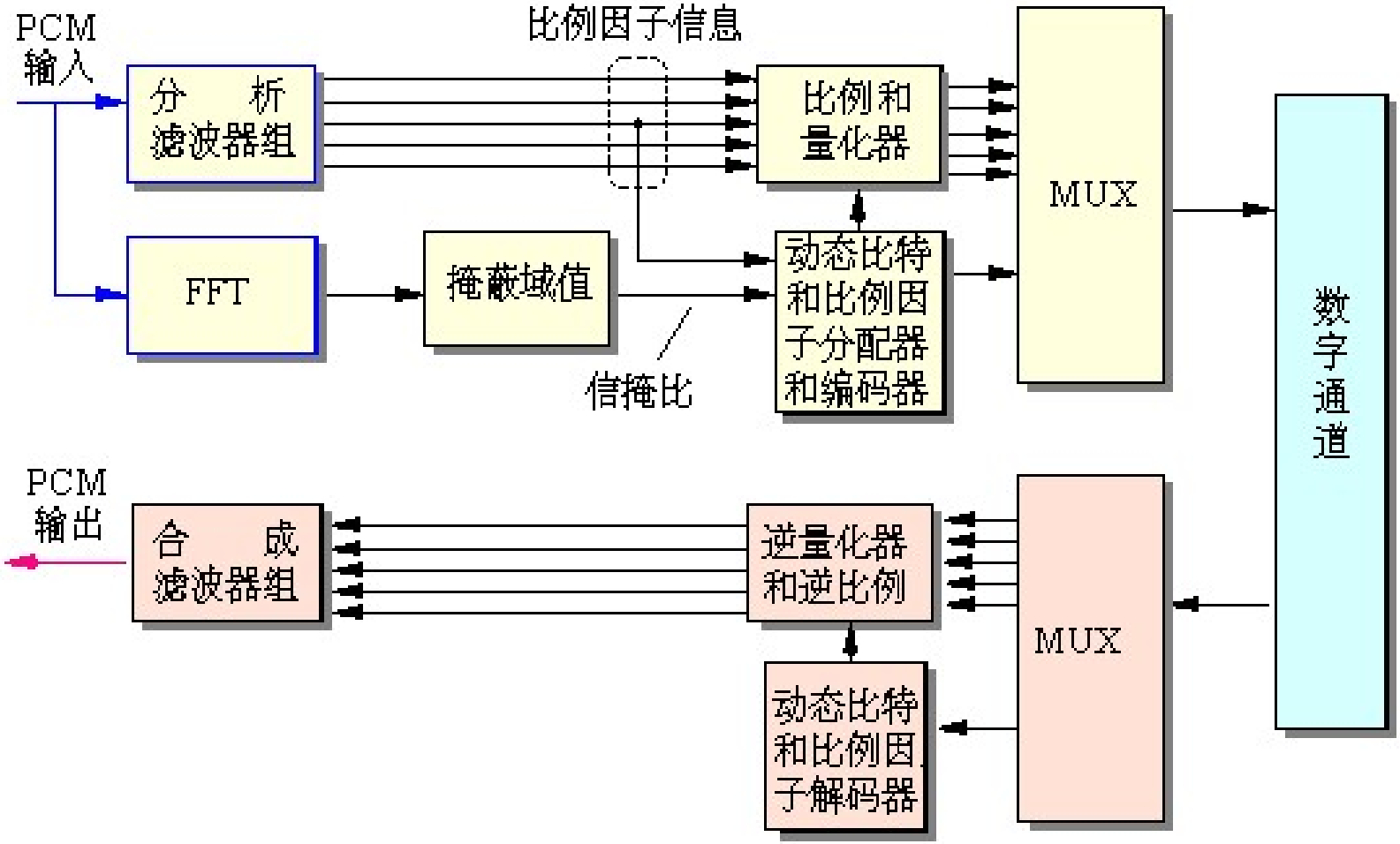
- **MPEG-1 ISO/IEC—11172-3**
  1993年标准化
- **MPEG-2 ISO/IEC—13818-3**
  1994年11月标准化,是对MPEG1的发展与扩展
- **ISO/IEC MPEG-2 AAC**
  (ISO/IEC 13818-7) 1997年4月公布
- **MPEG-4 ISO/IEC 14496-3**
  1999年标准化
- 美国**Dolby**实验室的**Dolby（AC-3）**
  1990年提出

四、 音频压缩的国际标准

- **MPEG-1 ISO/IEC—11172-3**
  1993年标准化
- **MPEG-2 ISO/IEC—13818-3**
  1994年11月标准化,是对MPEG1的发展与扩展
- **ISO/IEC MPEG-2 AAC**
  (ISO/IEC 13818-7) 1997年4月公布
- **MPEG-4 ISO/IEC 14496-3**
  1999年标准化
- 美国**Dolby**实验室的**Dolby（AC-3）**
  1990年提出

（一）、MPEG-1 音频压缩算法

MPEG-1 Audio层1和层2编解码器的结构



MPEG-1 Audio层1

1、滤波器组

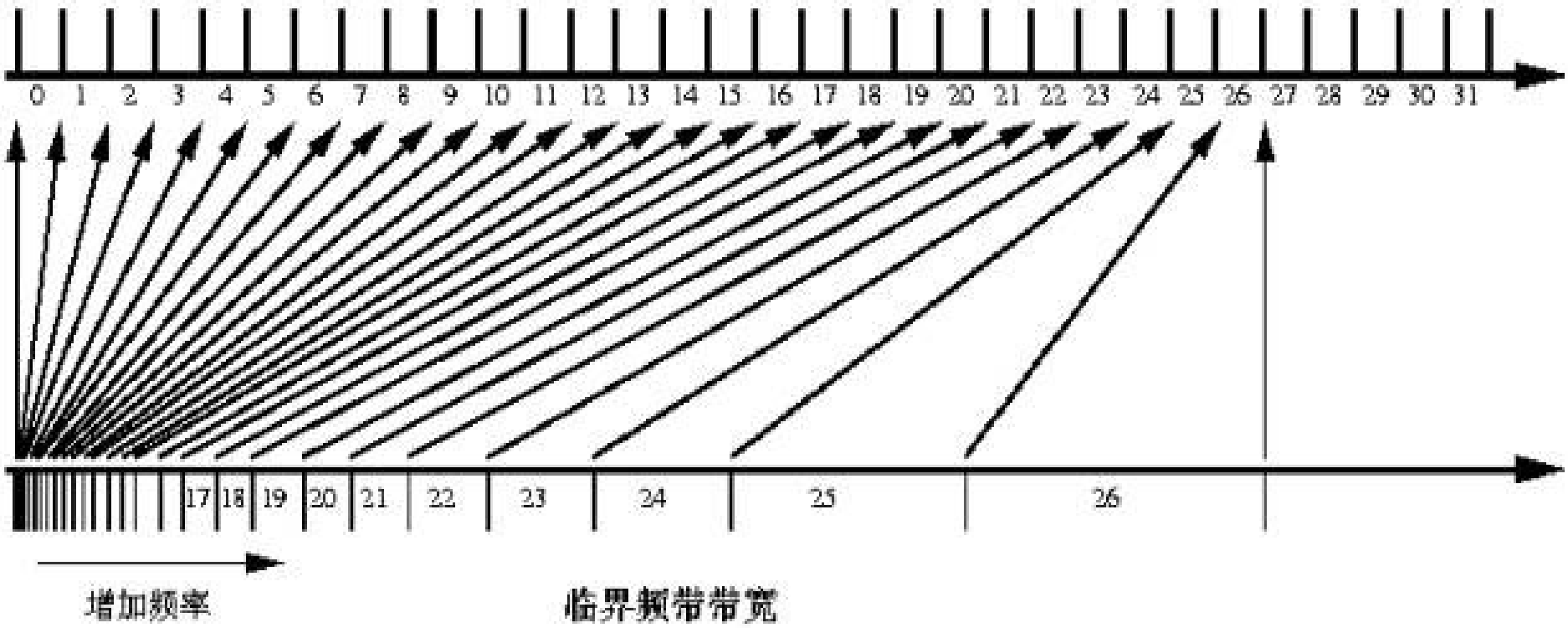
将时域信号变为32个等宽子带。

△f=（fs/2）/32=750Hz

最低频的子带滤波器为低通滤波器，其它为带通滤波器。

窄的子带能提高压缩比，改善声音质量.

MPEG/Audio 滤波器组频带



MPEG-1 Audio层1

2、快速傅利叶变(FFT）

作用：为满足掩蔽阈计算所需的精确的频谱分析，主要提高低频率范围的频率分辨率，与听觉特性相适应。

FFT的变换长度N＝512，取样频率*fs*＝48kHz时，通过FFT得到的频率分辨率为*fs*/512=93.75Hz

3、心理声学模型

模拟人耳听觉掩蔽特性的数学模型。

输入量：**FFT**的输出**X(K)**。

任务：计算信号掩蔽比**SMR**（每**8ms**计算**1**次）。

目的：根据**SMR**给各个子带分配量化级数（比特数）。

计算步骤：

(1)确定各子带的最大声级L(n)

(由12个连续抽样值的最大者确定)。

(2) 确定静听阈LTg。

(3) 确定音频信号中的音调（类似正弦信号） 成分和非音调（类似噪声）成分。

(4) 抽选掩蔽音，求出相关的掩蔽音。

(5) 计算相关掩蔽音各自的掩蔽阈(同听阈)。

(6) 计算总的掩蔽阈（同听阈）。

(7) 确定各子带中的最小掩蔽阈值（最小同听阈）

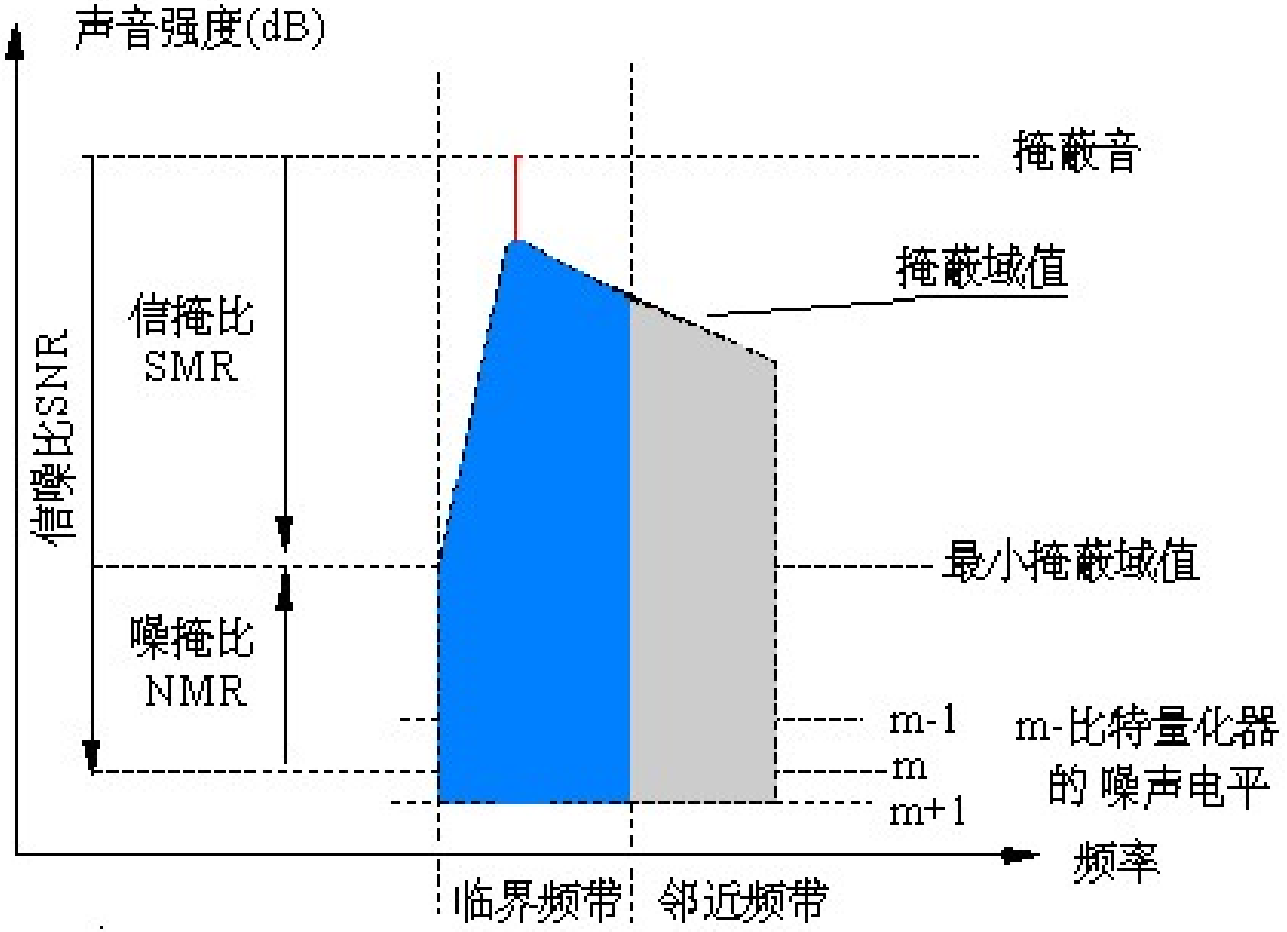
$LT_{min}(n)$。

(8) 计算各子带的信号掩蔽比

$SMR(n)=L(n)-LT_{min}(n)$。

信噪比（**SNR)=**信号峰值**-**噪声有效值（**dB**）

- 信号掩蔽比（**SMR)=** 信噪比**-**掩蔽噪声比（**dB**）



4、比例因子(SCF)

为了提高小信号的量化精度，不丢失小信号，对滤波器组输出的样值先进行归一化（如60dB)，大信号除以大于1的数，小信号除以 <1的数，这些除数即是比例因子，与音频数据一起传送，在解码端再恢复原有幅度。

（实际传送比例因子标记，查比例因子表可得因子）。

部分比例因子

标记iscf	比例因子	标记	比例因子iscf
0	2.00000000000000	56	0.00000480621738
1	1.58740105196820	57	0.00000381469727
2	1.25992104989487	58	0.00000302772723
3	1.00000000000000	59	0.00000240310869
4	0.79370052598410	60	0.00000190734863
5	0.62996052494744	61	0.00000151386361
7	0.50000000000000	62	0.00000120155435
…			

- 以块为单位记录一个因子：**12**个采样值，时间为**8ms**。
- 比例因子共**63**个，用**iscf=0**，**1**，**2……62**来标记，**6**比特字长编码。
- 例如，标记**iscf=0**的比例因子编码为“**000000**”，**iscf=62**的为“**111110**”。

5、动态比特分配

给每个子带分配多少比特进行量化，要同时满足比特率和掩蔽要求，总的原则是使音频帧期间的总的掩蔽噪声比达到最小。

动态的含义：

声音信号在不断随时间变化。

比特分配不是一次性完成，是一个迭代过程。

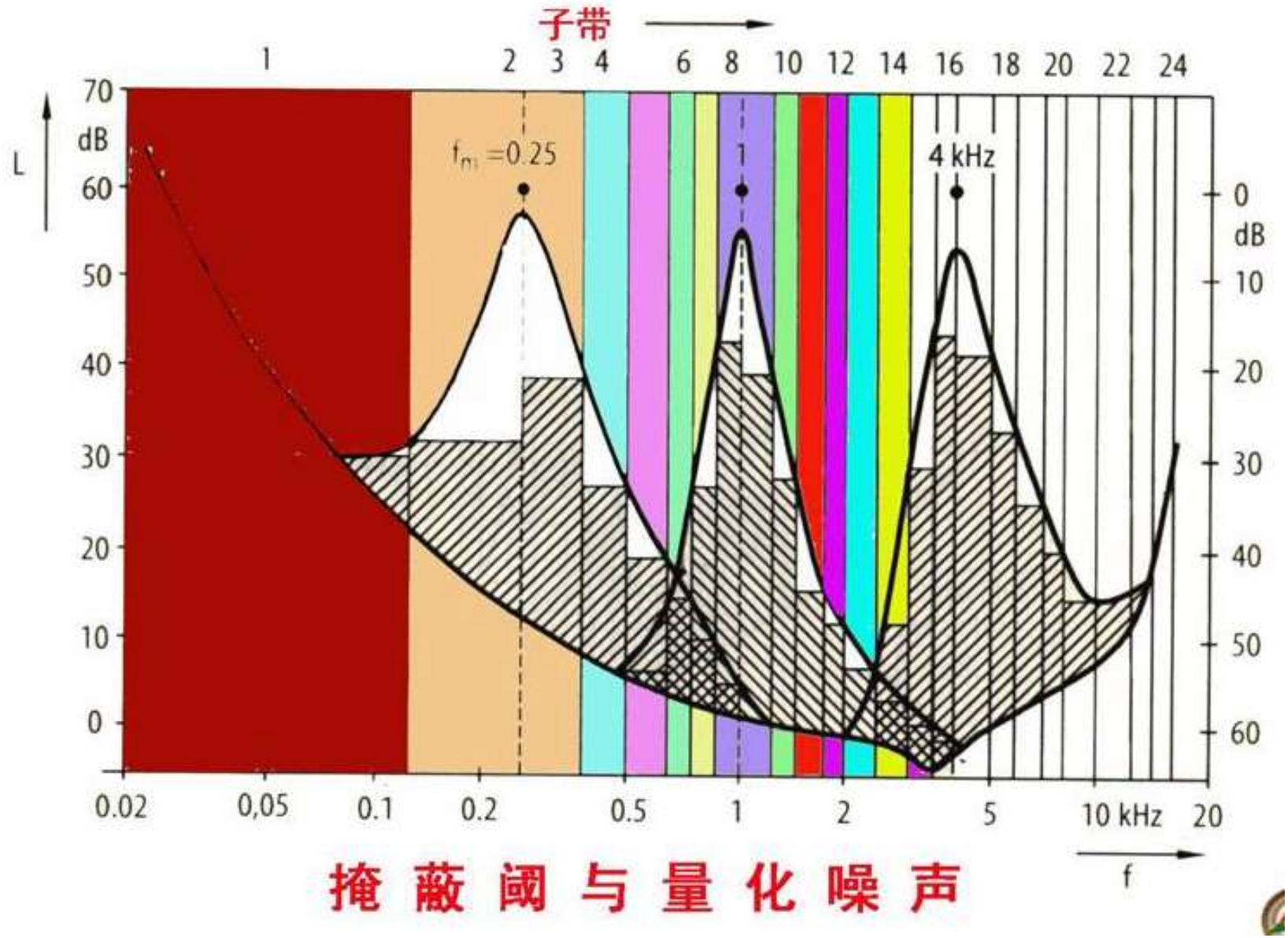
5、动态比特分配

“动态比特分配”：根据信号掩蔽比(SMR)确定子带的量化级数(比特数，对总数据率进行比特分配。

原则：（1）SMR(dB) = SNRmax(dB) – MNRmin(dB)

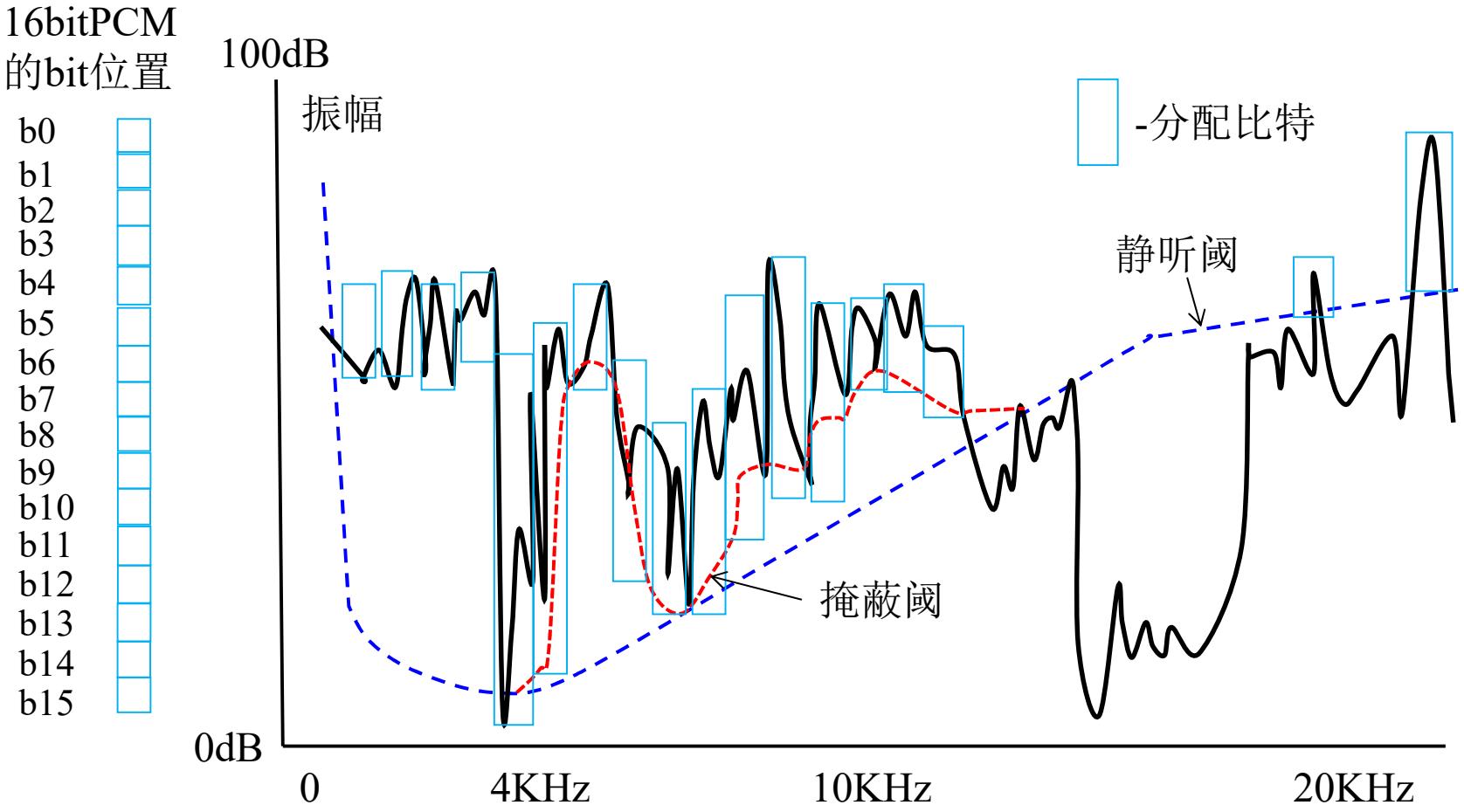
（2）使各子带的量化信噪比SNR>最小信掩蔽比SMR，将允许数据率分配给音频帧，再分给子带。音频帧的总的供使用的数据率扣除用于传送比例因子、比例因子选择信息、动态比特分配（BAL)、数据帧头与必要的差错检测和考虑附加数据后，分配给音频取样值。

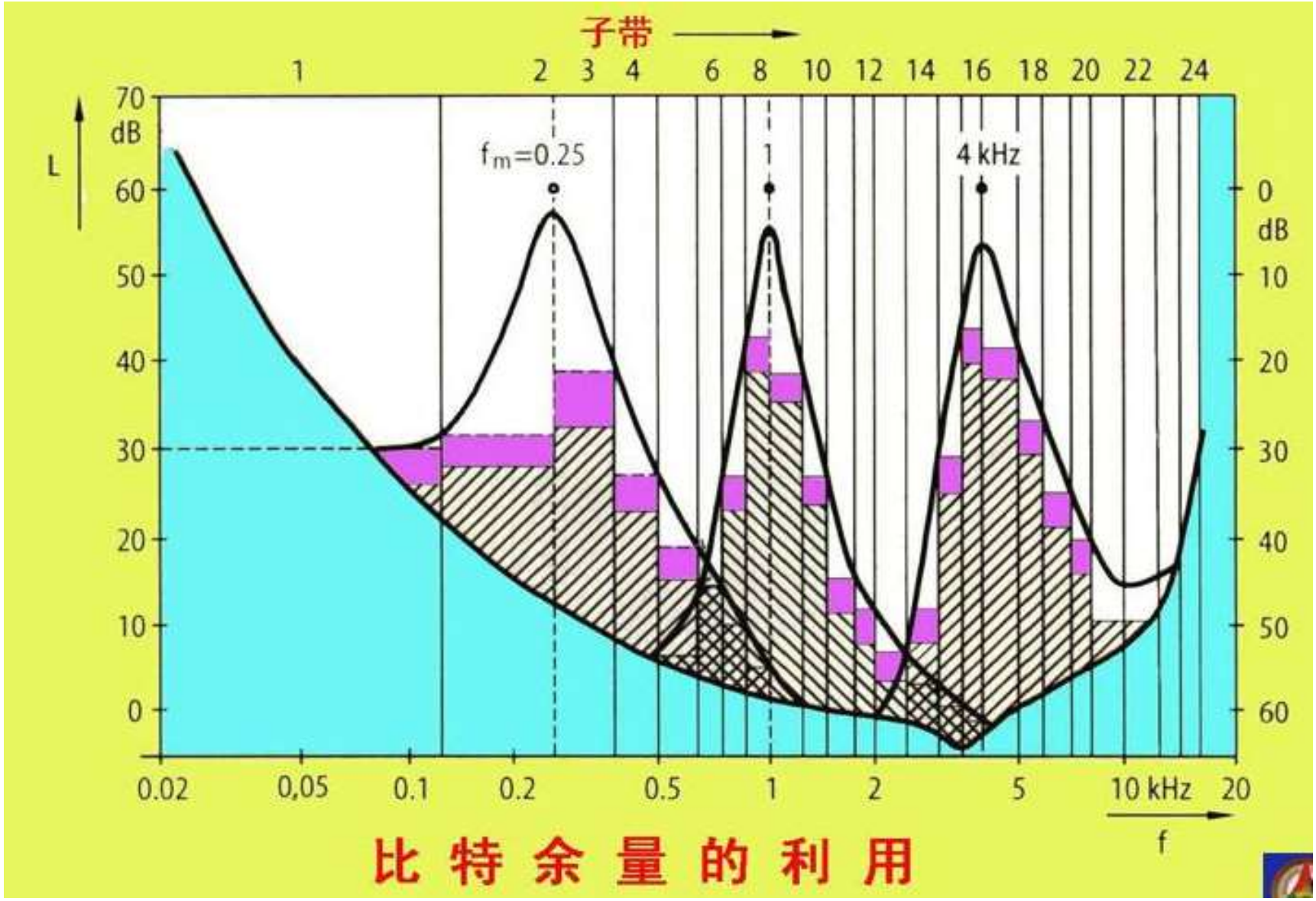
量化后　　S/N=6.02n+1.76 >SMR


子带
1
2
3
4
6
8
10
12
14
16
18
20
22
24
L
70
dB
60
50
40
30
20
10
0
f_{n_1} =0.25
1
4 kHz
0
dB
10
20
30
40
50
60
0.02
0,05
0.1
0.2
0.5
1
2
5
10 kHz
20
f

掩蔽阈与量化噪声

Example-动态比特分配




子带
1
2
3
4
6
8
10
12
14
16
18
20
22
24
L
70
dB
60
50
40
30
20
10
0
f_m=0.25
1
4 kHz
0
dB
10
20
30
40
50
60
0.02
0,05
0.1
0.2
0.5
1
2
5
10 kHz
20
f
比 特 余 量 的 利 用

6、子频带取样值的量化和编码

- 每子频带**12**个连续的样值都除以比例因子进行归一化，得到的值用**X**表示，进行量化计算：

A × X + B ，A和B：量化系数。

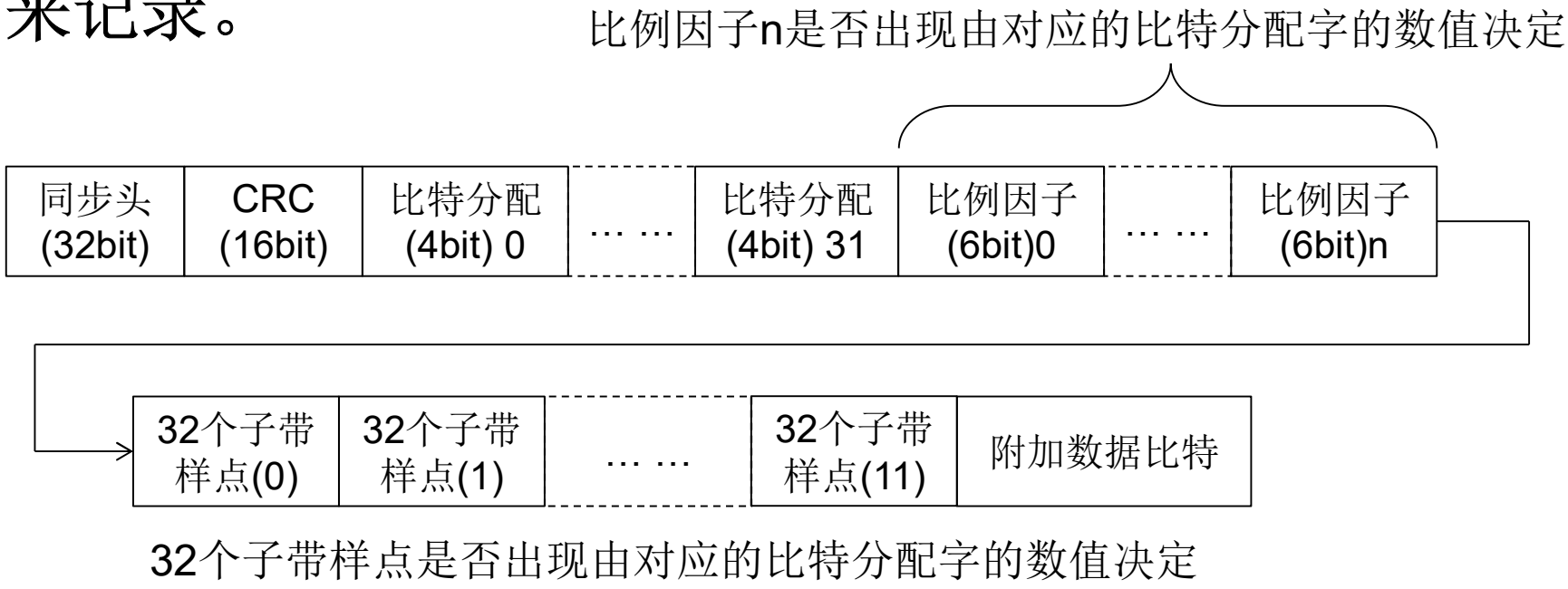
查量化表：根据**Bit**分配信息得量化级数，根据级数查量化表得**A**和**B**。

量化级	A	B	量化级	A	B
3	0.75	-0.25	1023	0.99902438	-0.000976563
7	0.625	-0.375	2047	0.999511719	-0.000488281
511	0.998046875	-0.001953125	65535	0.999984741	-0.000015259

7、帧结构

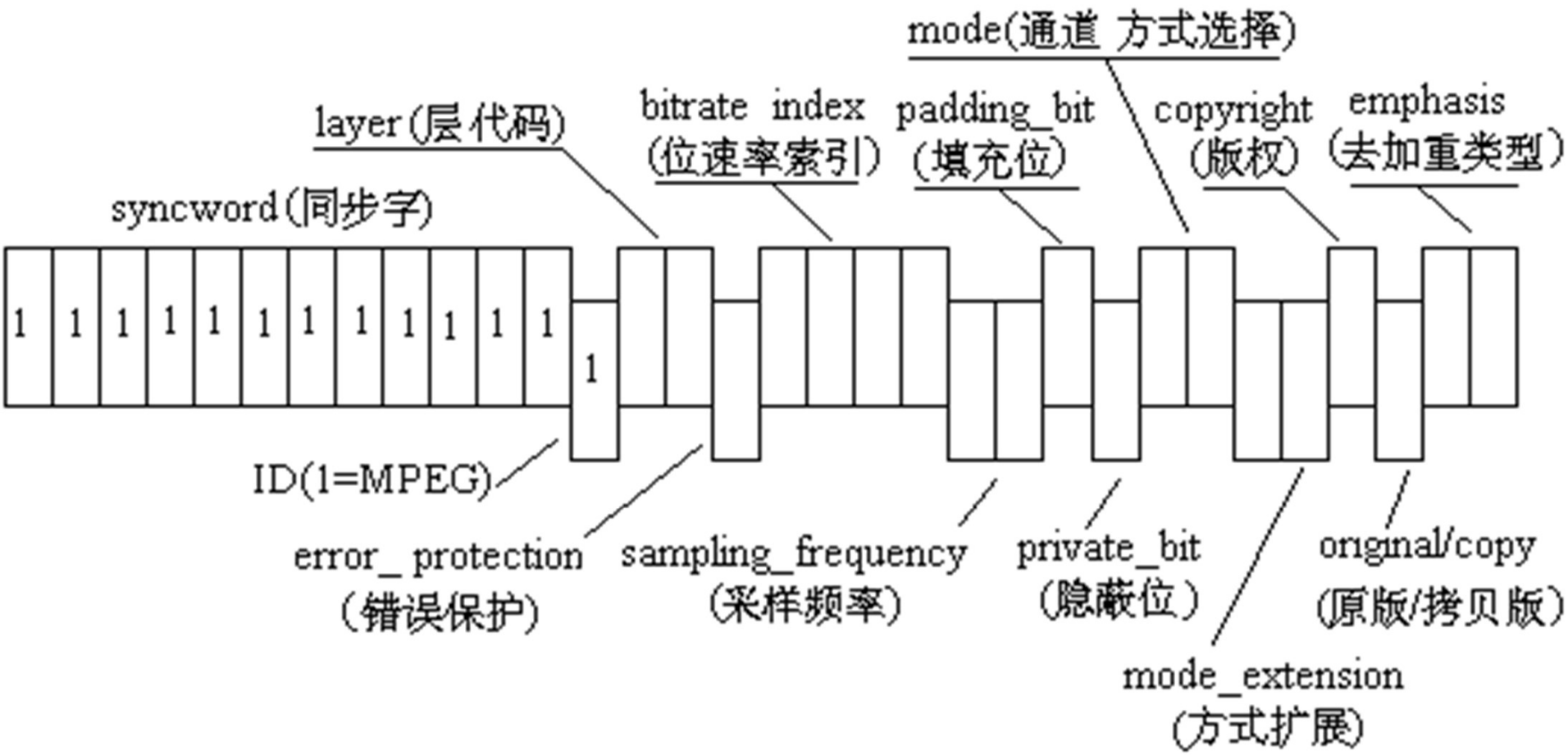
每个子带中前后相邻的连续12个样值合成一组，共用同一个比特分配值，以及同一个比例因子。

比特分配信息告诉解码器每个样本用多少比特来记录。



- audio_data()
- {
- if (mode==single_channel)
- {
- for (sb=0; sb<32; sb++)
- allocation[sb] 4 bits uimsbf
- for (sb=0; sb<32; sb++)
- if (allocation[sb]!=0)
- scalefactor[sb] 6 bits uimsbf
- for (s=0; s<12; s++)
- for (sb=0; sb<32; sb++)
- if (allocation[sb]!=0)
- sample[sb][s] 2..15 bits uimsbf
- }……}

MPEG声音比特流同步头的格式



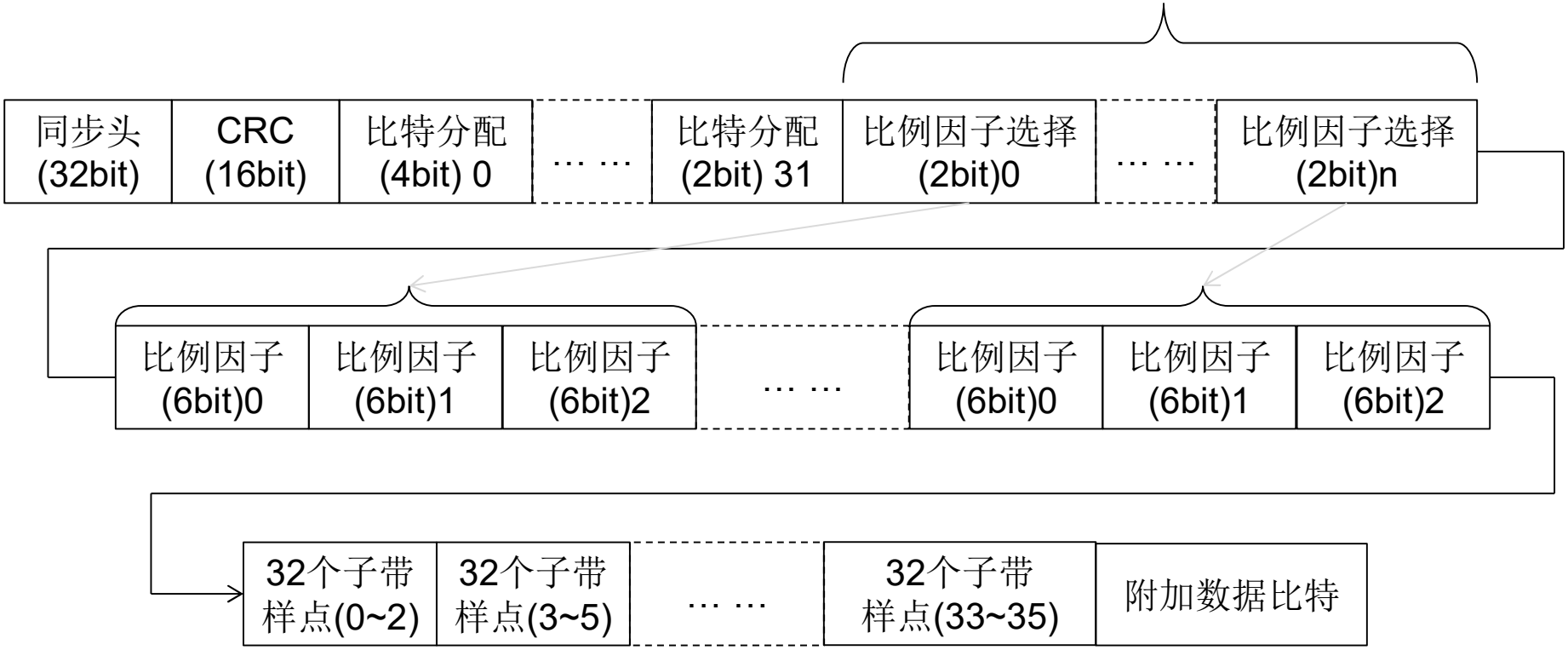
MPEG-1 Audio层2

层2与层1的不同之处：

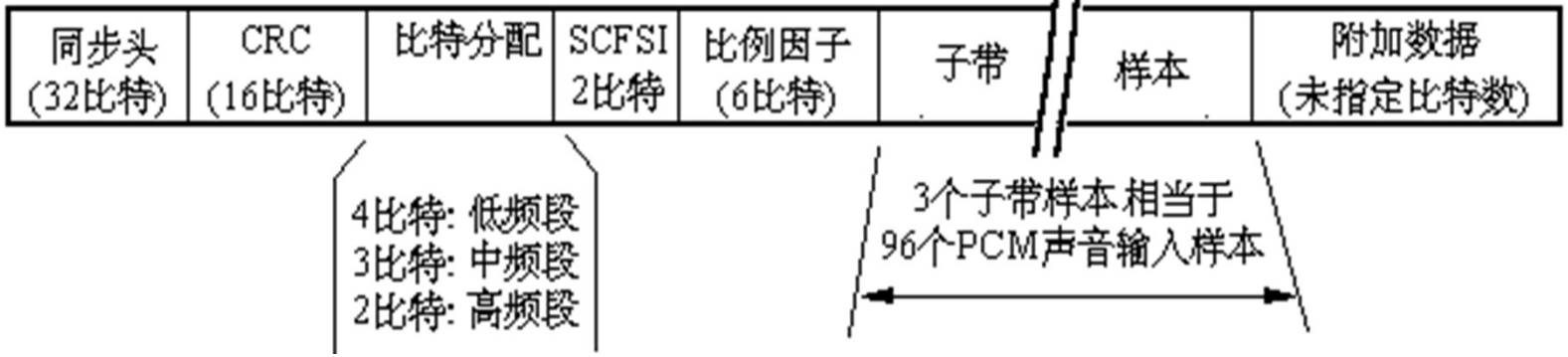
- 使用1024点的FFT运算，提高频率分辨率。
- 抽样频率为48KHz时，音频帧长24ms，含样值数1152个，每个子带含1152/32=36个，鉴于人耳听觉的时间掩蔽特性，每12个抽样值归并成一个块，3个块，时间：12 ×32/48＝8ms.
- 描述比特分配的字段随子带不同而不同。
- 一个子带3组样值使用三个不同比例因子，传送比例因子选择信息（SCFSI）.

MPEG音频层 2 的帧结构

比例因子选择n是否出现由对应的比特分配字的数值决定



32个自带样点是否出现由对应的比特分配字的数值决定



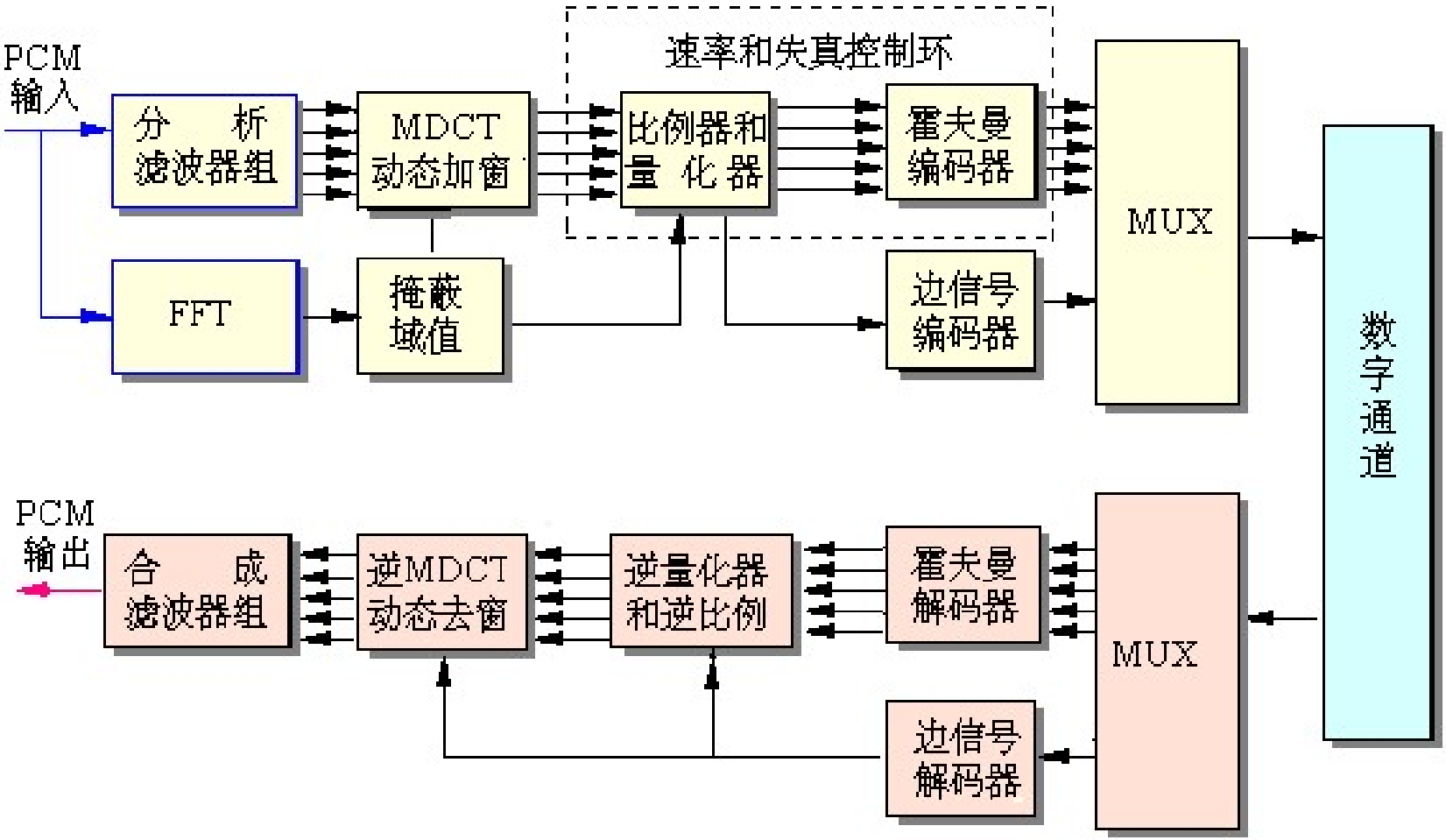
```
audio_data()
{  if (mode==single_channel)
 {   for (sb=0; sb<sblimit; sb++)
 allocation[sb]                          2..4 bits uimsbf
 for (sb=0; sb<sblimit; sb++)
 if (allocation[sb]!=0)
 scfsi[sb]                               2 bits bslbf
 for (sb=0; sb<sblimit; sb++)
 if (allocation[sb]!=0)
 {  if (scfsi[sb]==0)
 { scalefactor[sb][0]                    6 bits uimsbf
 scalefactor[sb][1]                      6 bits uimsbf
 scalefactor[sb][2] }                    6 bits uimsbf
 if (scfsi[sb]==1) || (scfsi[sb]==3)
 { scalefactor[sb][0]                    6 bits  uimsbf
 scalefactor[sb][2] }                    6 bits  uimsbf
 if (scfsi[sb]==2)
 scalefactor[sb][0]                      6 bits  uimsbf
 }
 for (gr=0; gr<12; gr++)
 for (sb=0; sb<sblimit; sb++)
 if (allocation[sb]!=0)
 {
 if (grouping[sb])
 samplecode[sb][gr]     5..10 bits  uimsbf
 else for (s=0; s<3; s++)
 sample[sb][3*gr+s]     2..16 bits  uimsbf
 }
 } …….}
```

MPEG-1 音频压缩算法

- 层1的子带是频带相等的子带，它的心理声学模型仅使用频域掩蔽特性,FFT为512点。
- 层2对层1作了改进，相当于3个层1的帧，每帧有1152个样本。它使用的心理声学模型利用频域及时间掩蔽特性，并且在低、中和高频段对比特分配作了一些限制，对比特分配、比例因子和量化样本值的编码也更紧凑。
- 层3把声音频带分成非等带宽的子带，心理声学模型除了使用频域掩蔽特性和时间掩蔽特性之外，还考虑了立体声数据的冗余。

MPEG-1 Audio层3

编解码器的结构



MPEG-1 Audio层3

- 采用**MDCT (**改进离散余弦变换**)**，消除混叠效应。指定**2**种**MDCT**块长，长块为**18**个样本，短块为**6**个样本，相邻窗口之间有**50**％的重叠。

  长块对于平稳的音频信号提高频域分辨率；

  短块对瞬变的音频信号提高时域分辨率。一般对最低频的两个子带使用长块，其余使用短块。
- 对量化值用**VLC**，即霍夫曼**(Huffman)**编码。总数据率**32~320kbps**，节省**20**％的码率。
- 代价：编码器的复杂度和解码器缓冲容量。

课程小结

- 听觉系统特性
  - 人耳的带通滤波器组
  - 掩蔽效应: 静听域, 频域掩蔽/时域掩蔽
- 音频感知编码基本组成
  - 子带分割、基于心里学模型的比特分配、量化、编码
- **MPEG1** 音频
  - **Layer1、2、3**的特点和性能
  - 三个层次编码的不同点

作业7（2020年4月21日）

- 1、 对声音信号进行压缩编码的依据是什么？
- 2、 什么是临界频带？什么是频域掩蔽？什么是时域掩蔽？
- 3、 子带编码的基本思想是什么？使用子带编码的好处？
- 4、在MPEG-1音频压缩编码中，对声音信号进行量化使用比例因子有什么作用？
- 5、在MPEG-1音频压缩编码中，为什么采用动态比特分配？
- 6 、简述MPEG-1AUDIO layer1,2,3各自编码原理有什么特点？分别应用于哪些场合？