



第七章 分析-综合编码

7.1 引言

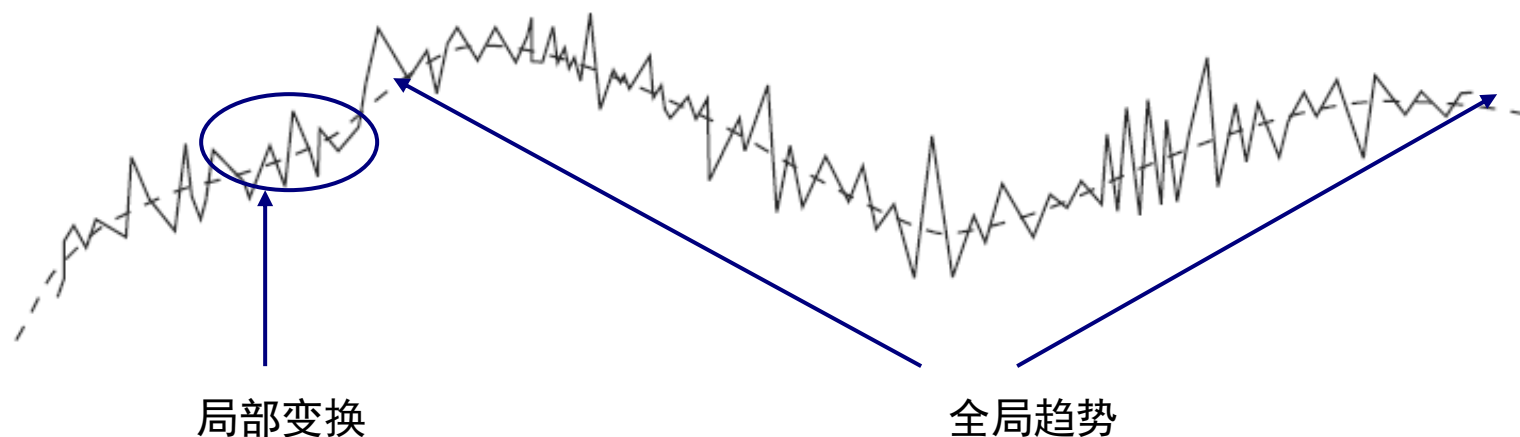
- 基于块变换的编码将信源输出分解为不同频率子带，然后对不同频率的子带编码
 - 但当码率很低时，会表现出严重的块效应 → 可用LOT (lapped orthogonal transform)避免块效应
- 子带编码：将原始信号分解为若干个子频带，对其分别进行编码处理后再合成为全频带信号
- 小波变换编码：使用更强大的非均匀分辨率滤波器对信号进行时间-频率局部分分析与综合

动机

■ 观察

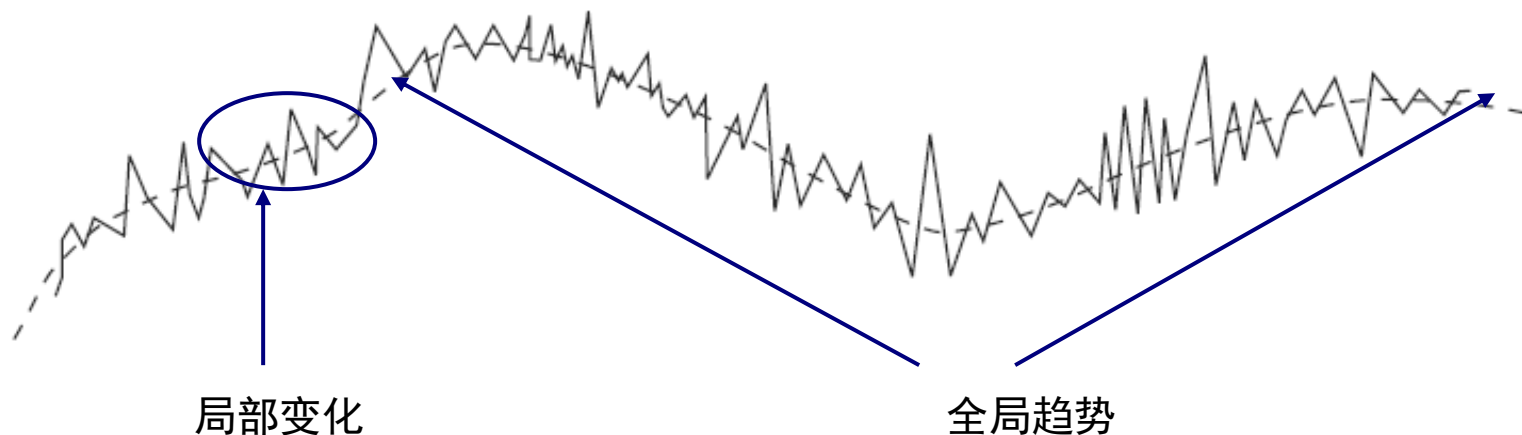
- 很多信源在同一时间表现出不同的特性
 - 不同模型
 - 人对不同的失真的感知敏感度不同
- ➔ 没有一个单独的模型能满足所有的情况

■ 例：



快速变化的信源输出,其中含有长期缓慢变化的分量

7.2.1 信号的分解与合成



提取全局趋势：在一个滑动窗口上计算样本均值

$$y_n = \frac{1}{2}(x_n + x_{n-1})$$

■ $\{y_n\}$ 比 $\{x_n\}$ 更适合用差分编码方式

■ 为了得到 x_n ，还需计算 z_n ：

$$z_n = x_n - y_n = x_n - \frac{1}{2}(x_n + x_{n-1}) = \frac{1}{2}(x_n - x_{n-1})$$

■ 对 $\{y_n\}$ 和 $\{z_n\}$ 可以采用不同的编码方式

7.2.1 信号的分解与合成

例7.1:

$\{x_n\}$: 10 14 10 12 14 8 14 12 10 8 10 12

$\{x_n - x_{n-1}\}$: 10 4 -4 2 2 -6 6 -2 -2 -2 2 2

Sol #1: M -水平量化 (忽略第一个值)

$$M = 2^m$$

$$\Delta = 12/M$$

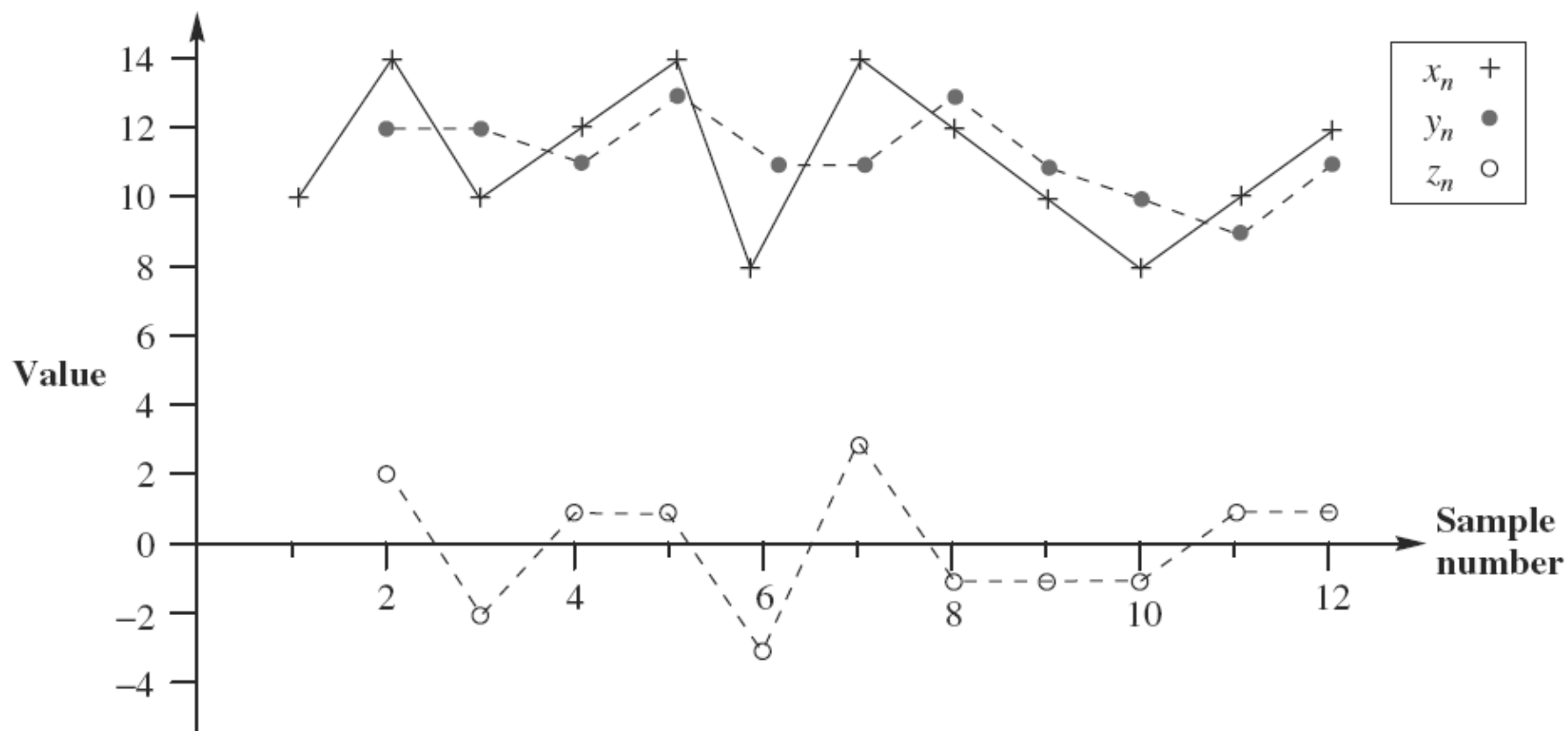
$$\text{最大量化误差}(\text{MaxQE}) = \Delta/2 = 6/M$$

Sol #2: 利用 $\{y_n\}$ $\{z_n\}$

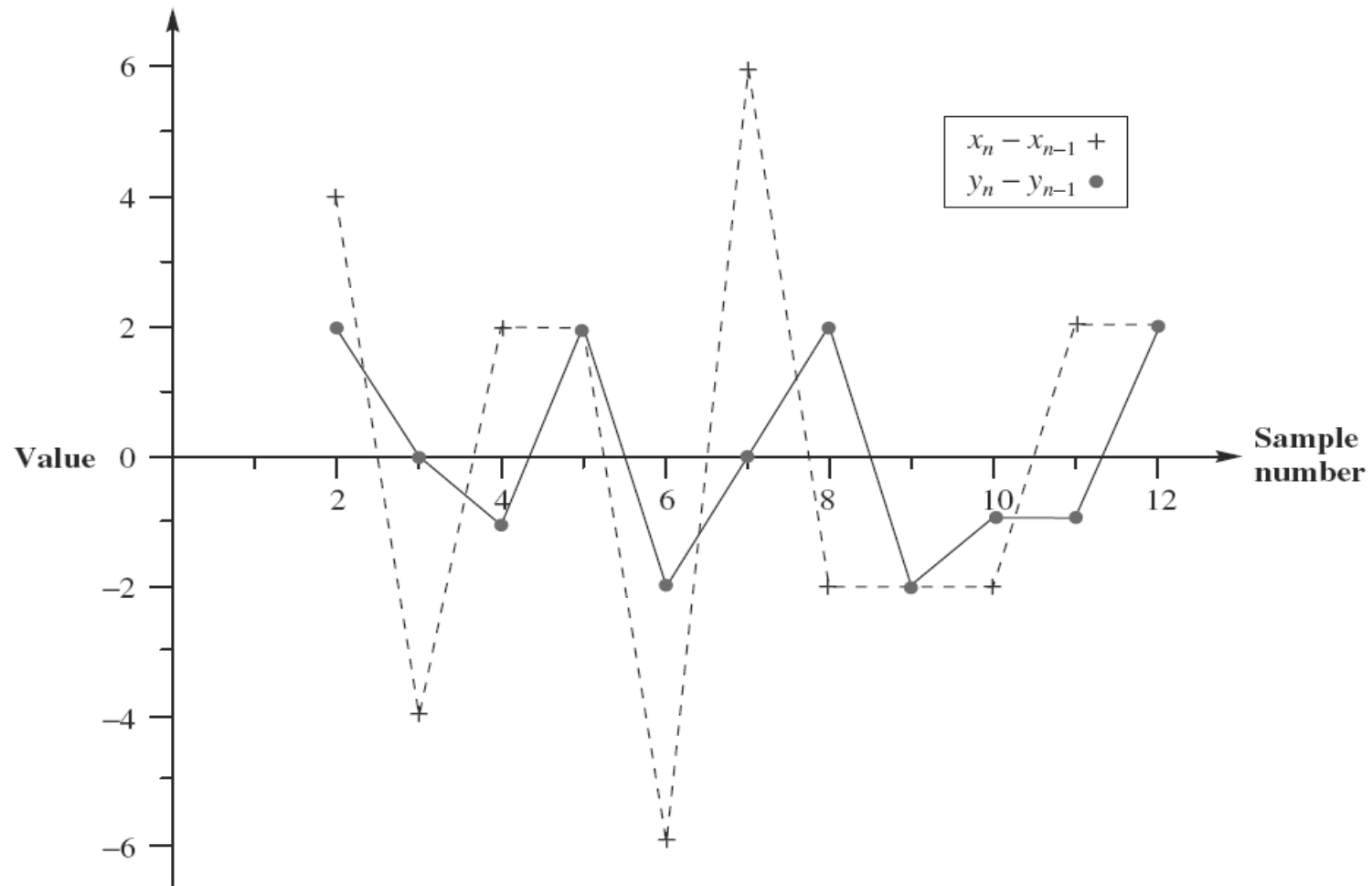
$\{y_n\}$: 10 12 12 11 13 11 11 13 11 10 9 11

$\{y_n - y_{n-1}\}$: 10 2 0 -1 2 -2 0 2 -2 -1 -1 2

7.2.1 信号的分解与合成



7.2.1 信号的分解与合成



7.2.1 信号的分解与合成

$\{y_n - y_{n-1}\}$ 的动态范围为4

M-水平量化器

$$\Delta = 4/M$$

$$\text{MaxQE} = 2/M$$

$\{z_n\}$:

0 2 -2 1 1 -3 3 -1 -1 -1 1 1

动态范围 = 6

$$\Delta = 6/M$$

$$\text{MaxQE} = 3/M$$

至此为止

编码了**2倍**的数值，所以比特率也变成了2倍

失真更小：5/M

7.2.1 信号的分解与合成

降低比特率

将 $\{y_n\}$ 分解成 $\{y_{2n}\}$ & $\{y_{2n-1}\}$

将 $\{z_n\}$ 分解成 $\{z_{2n}\}$ & $\{z_{2n-1}\}$

只需传送偶数下标的子序列（或奇数下标子序列）

$$y_{2n} = \frac{1}{2}(x_{2n} + x_{2n-1})$$

$$y_{2n} + z_{2n} = x_{2n}$$

$$z_{2n} = \frac{1}{2}(x_{2n} - x_{2n-1})$$

$$y_{2n} - z_{2n} = x_{2n-1}$$

- 我们又回到了原始比特率，量化误差仍为 $5/M$

7.2.1 信号的分解与合成

在两种情况下，最后传送数值的数目相同(n)

两个子序列有不同的特征

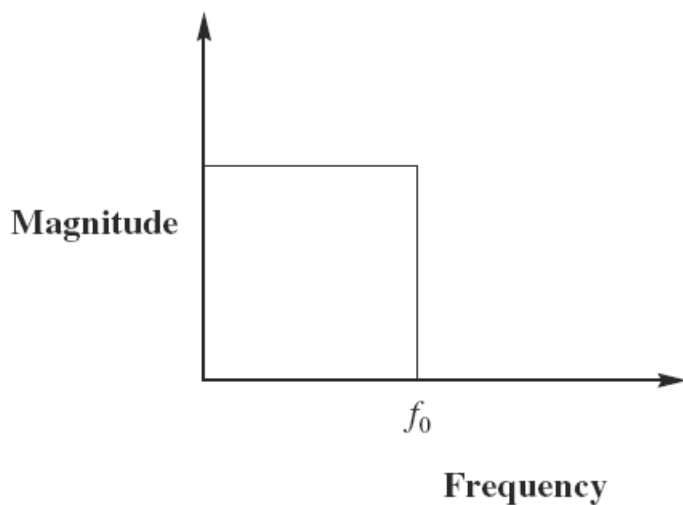
因此可以采用不同的编码机制
编码更灵活，从而编码效率更高

还可以递归使用该分解方法，得到子序列
亦称为分析—综合(*analysis & synthesis*)

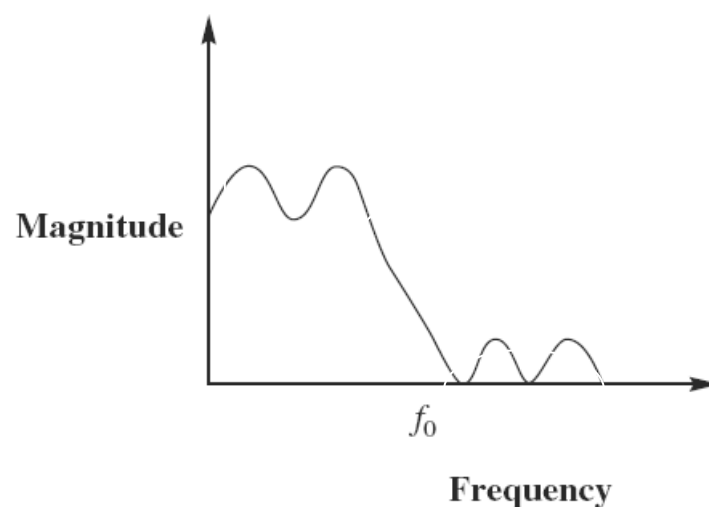
信号分解可利用数字滤波器实现

滤波器

- 滤波器：分离特定频率成分的系统
- 幅值传递函数(Magnitude transfer function):
 - 滤波器输入与输出幅值的比率，为频率的函数



理想低通滤波器



实际低通滤波器

- 实际低通滤波器的截至更缓慢，且在带通区域和截至区域很产生波纹(ripple)

7.2.2 滤波器组的设计

■ 数字滤波器回顾

- 当前和过去输入（和输出）的加权组合

$$y_n = \sum_{i=0}^N a_i x_{n-i} + \sum_{i=1}^M b_i y_{n-i}$$

滤波器系数

■ 脉冲：

- $\{x_n\} = 1\ 0\ 0\ 0\ \dots$

■ 脉冲响应

- 有限 → 有限脉冲响应(Finite Impulse Response, FIR)
- 无限 → 无限脉冲响应(Infinite Impulse Response, IIR)
- 注意： $b_i = 0 \rightarrow \text{FIR}$

7.2.2 滤波器组的设计

■ **例7.2** $a_0 = 1.25, a_1 = 0.5$

$$x_n = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases} \quad \begin{aligned} y_0 &= a_0 x_0 + a_1 x_{-1} = 1.25 \\ y_1 &= a_0 x_1 + a_1 x_0 = 0.5 \\ y_n &= 0, \quad n < 0 \text{ or } n > 1 \end{aligned}$$

■ 脉冲响应 $\{h_n\}$

$$h_n = \begin{cases} 1.25 & n = 0 \\ 0.5 & n = 1 \\ 0 & \text{otherwise} \end{cases}$$

■ 注意 $\{h_i\} \rightarrow \{a_i\}$

■ 脉冲响应函数完全决定滤波器

7.2.2 滤波器组的设计

■ 例7.3 如例7.1中

$$y_n = \frac{1}{2}(x_n + x_{n-1})$$

$$z_n = \frac{1}{2}(x_n - x_{n-1})$$

■ $\{y_n\}$ $\{z_n\}$ 序列为2抽头(tap) FIR滤波器:

$$h_n = \begin{cases} 0.5 & n = 0 \\ 0.5 & n = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$h_n = \begin{cases} 0.5 & n = 0 \\ -0.5 & n = 1 \\ 0 & \text{otherwise} \end{cases}$$

7.2.2 滤波器组的设计

■ 例7.4 令 $a_0 = 1, b_1 = 0.9$

$$y_0 = a_0 x_0 + b_1 y_{-1} = 1(1) + 0.9(0) = 1$$

$$y_1 = a_0 x_1 + b_1 y_0 = 1(0) + 0.9(1) = 0.9$$

$$y_2 = a_0 x_2 + b_1 y_1 = 1(0) + 0.9(0.9) = 0.9^2$$

\vdots

$$y_n = (0.9)^n$$

$$h_n = \begin{cases} 0 & n < 0 \\ (0.9)^n & n \geq 0 \end{cases}$$

■ 注意：这是一个**IIR**滤波器

7.2.2 滤波器组的设计

- 脉冲响应函数完全规定了一个FIR & IIR滤波器：

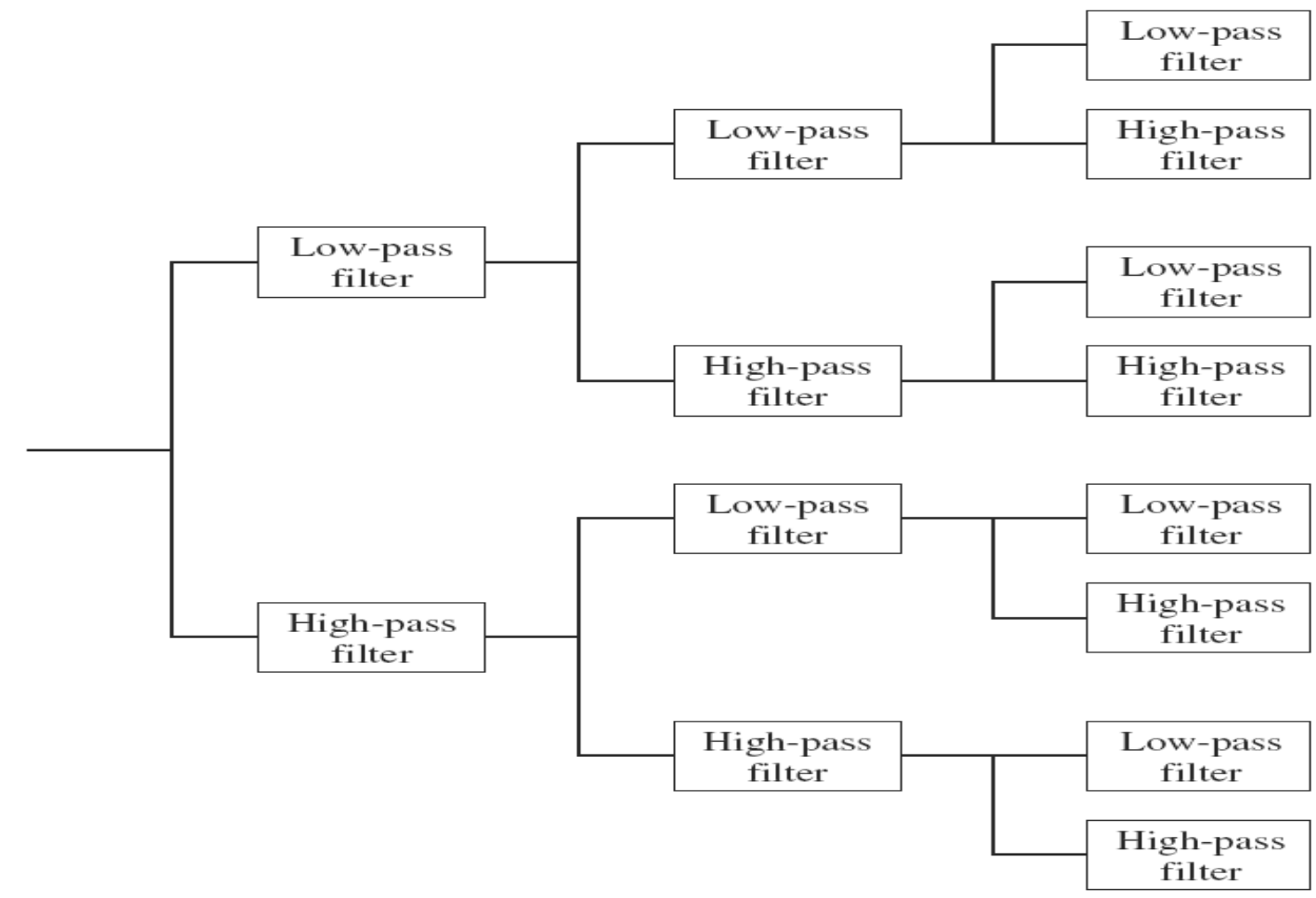
$$y_n = \sum_{k=0}^M h_k x_{n-k}, (M < \infty \text{ for FIR}, M = \infty \text{ for IIR})$$

↑
卷积

- 平稳性

- 如果一个滤波器的输入是有限的意味着其输出也是有限的，则该滤波器是平稳的
- FIR滤波器通常是平稳的：加权平均
- IIR可能对有界的输入产生无界的输出。如：
 - $a_0 = 1, b_1 = 2 \rightarrow y_n = 2^n$
 - 虽然IIR滤波器可能不稳定，但能提供更快的截止和产生更少ripple

7.2.2 滤波器组的设计



8-带树状滤波器组

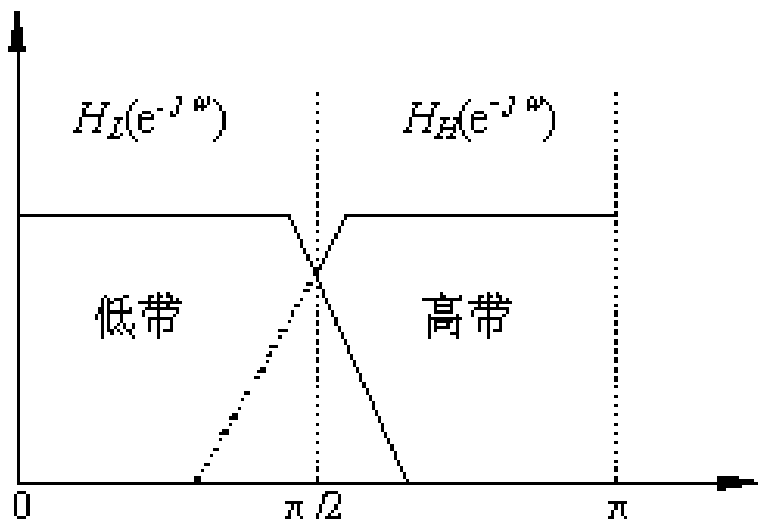
7.2.2 滤波器组的设计

■ 最常用的滤波器组

- 正交镜像滤波器(Quadrature mirror filters, QMF)

■ 基本性质

- 如果低通滤波器由 $\{h_n\}$ 给定,
则高通滤波器由 $\{(-1)^n h_{N-1-n}\}$ 给定
- QMF幅频特性简图



7.2.2 滤波器组的设计

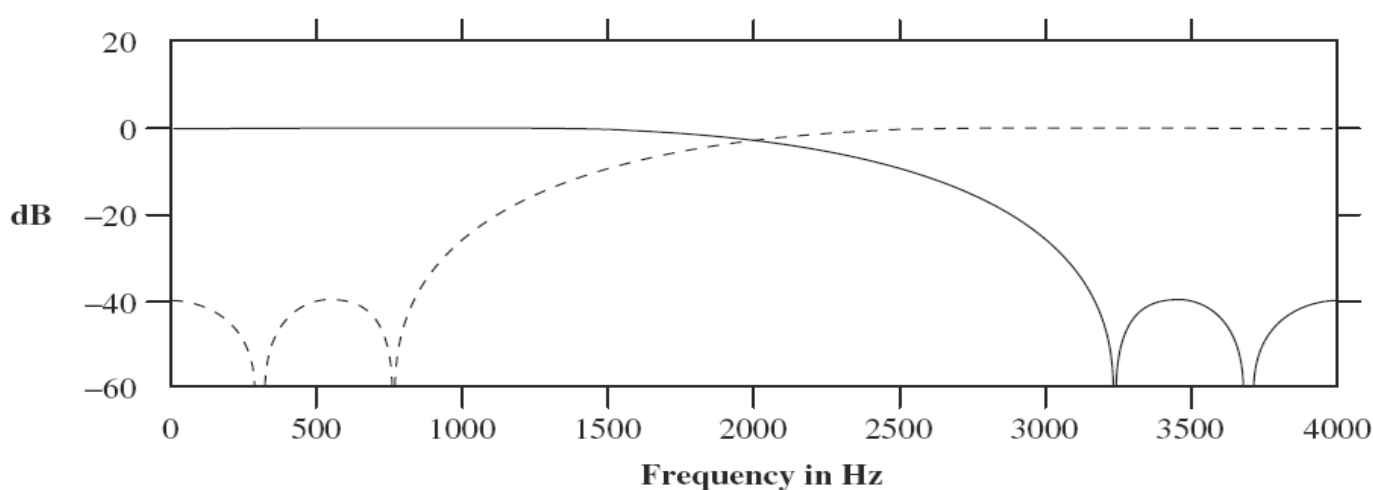
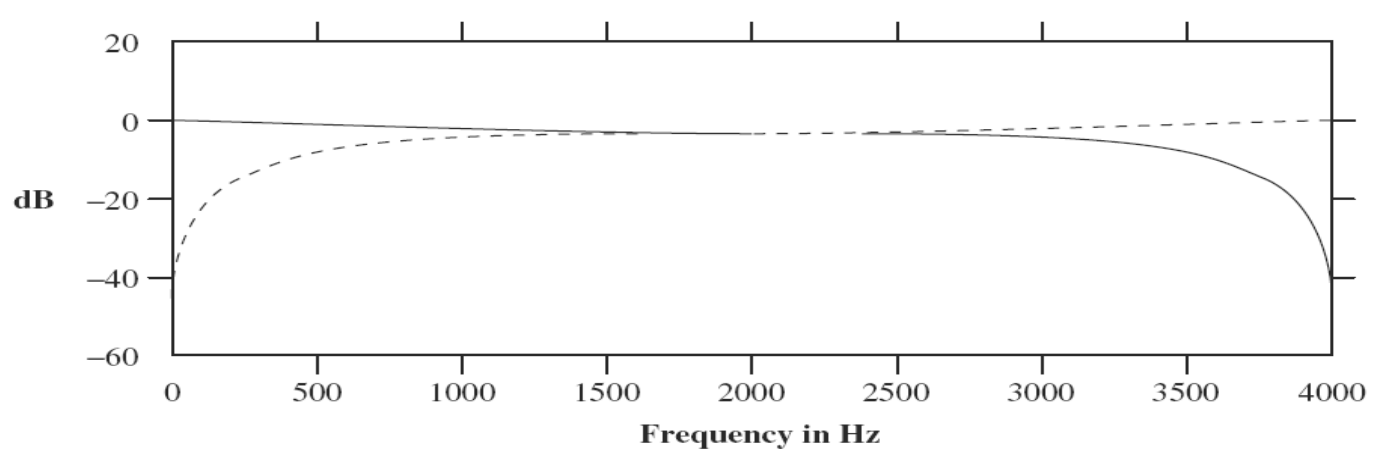
Johnston QMF滤波器

h_0, h_7	0.00938715	h_0, h_{15}	0.002898163
h_1, h_6	0.06942827	h_1, h_{14}	-0.009972252
h_2, h_5	-0.07065183	h_2, h_{13}	-0.001920936
h_3, h_4	0.48998080	h_3, h_{12}	0.03596853
		h_4, h_{11}	-0.01611869
		h_5, h_{10}	-0.09530234
		h_6, h_9	0.1067987
		h_7, h_8	0.4773469

- 抽头更多，分解效率更高，但计算更复杂

$$y_n = \sum_{k=0}^M h_k x_{n-k}$$

Johnston vs. Smith-Barnwell



Smith-Barnwell滤波器的截止更快→ 分离效果更好

7.2.2 滤波器组的设计-余弦调制QMF

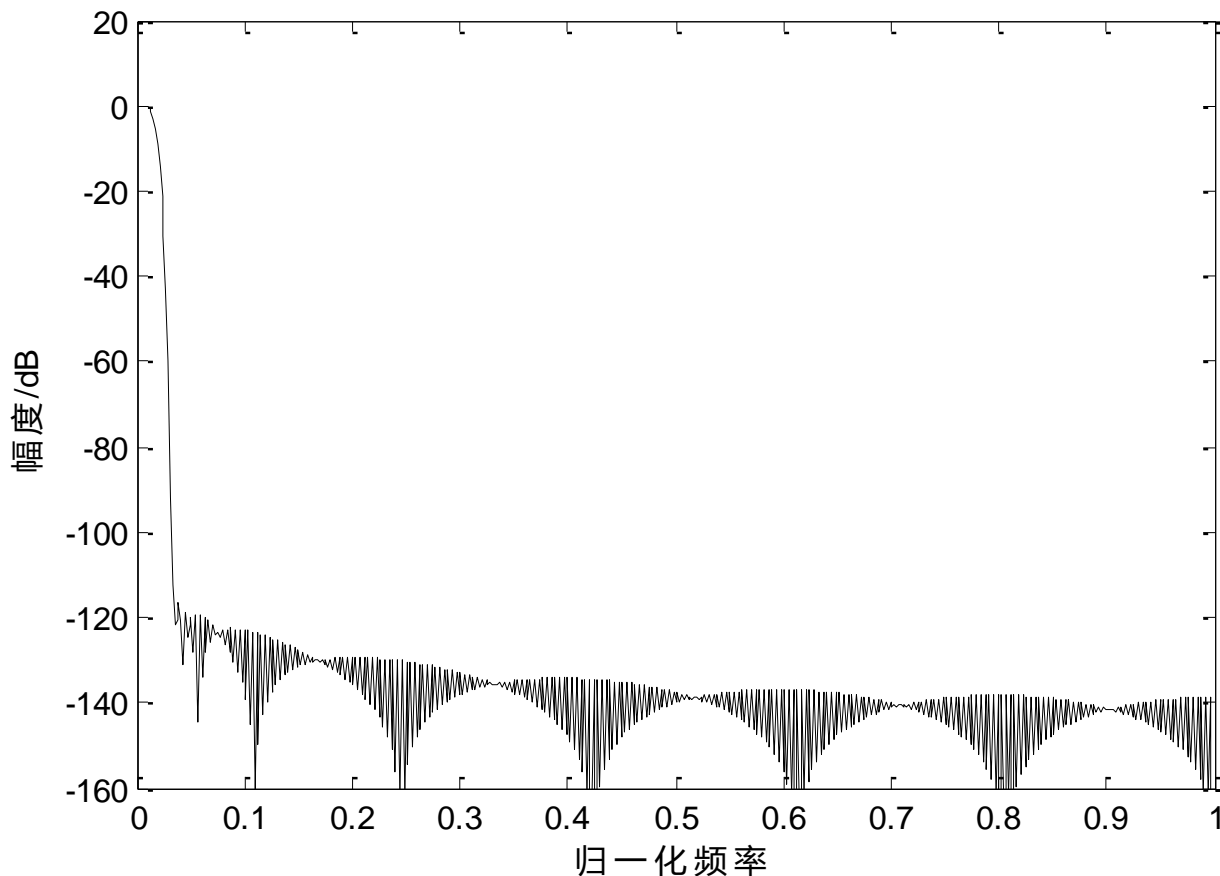
- 多相滤波器组(Polyphase Filter Bank): 将PCM样本变换到32个子带的频域信号
 - 带通滤波器系数相当于用该子带中心频率的正弦分量与 $\{h(i)\}$ 相乘。可见每个子带滤波器组相当于一低通滤波器 $\{h(i)\}$ 与各子带中心频率的余弦函数（对称！）相乘(调制)结果，而且通过设计低通滤波器特性和调整各余弦函数相位，

$$h(i) \times \cos[(2n-1)\pi/2M - \phi_n] \quad i=0,1,\dots,511, \quad n=1,2,\dots,32$$

其中 ϕ_n 为相位， $M=32$ 。BP_n 的输出为

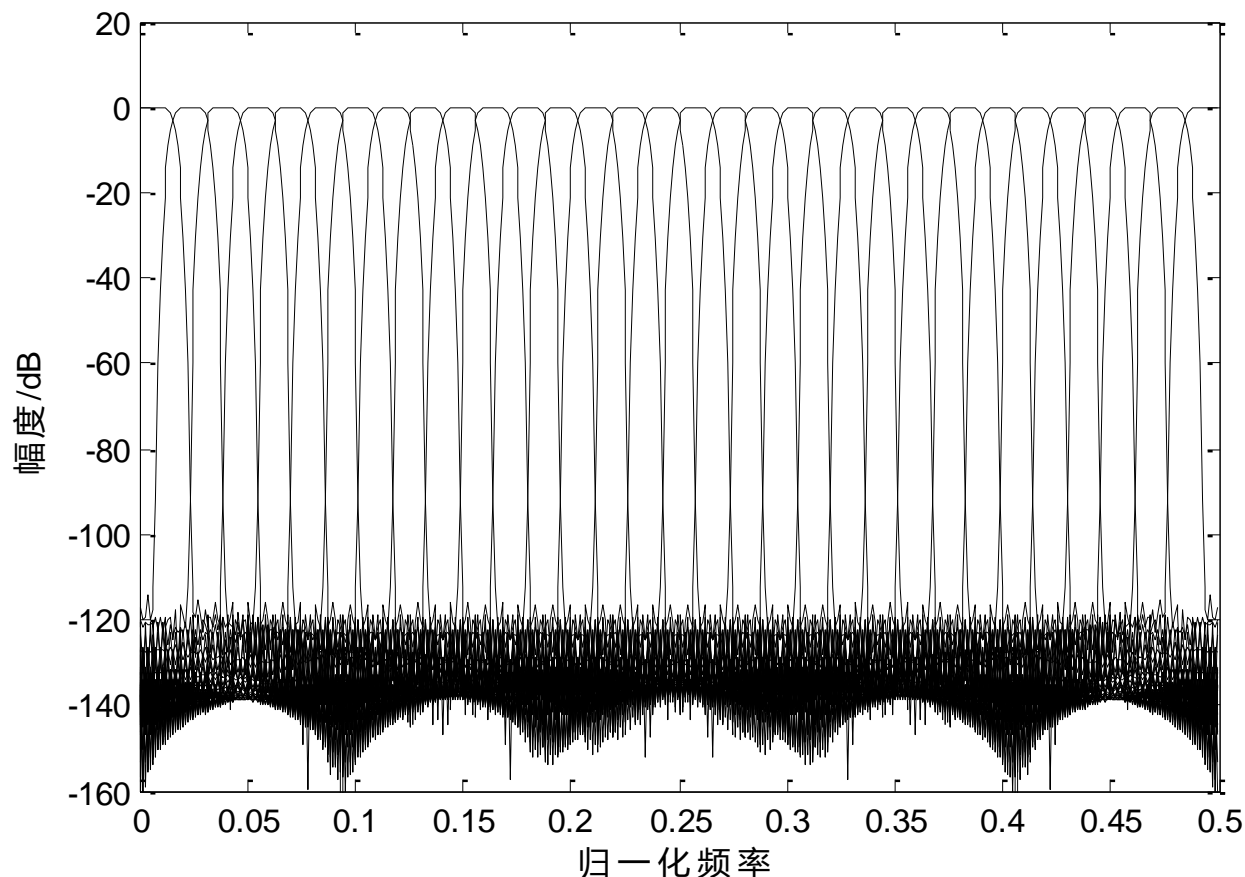
$$Y_{bp_n}(k) = \sum_{i=0}^{511} x(k-i)h(i)\cos[(2n-1)i\pi/2M - \phi_n] \quad n=1,2,\dots,32$$

7.2.2 滤波器组的设计-余弦调制QMF



原型滤波器幅频响应示意图

7.2.2 滤波器组的设计-余弦调制QMF



分析滤波器组幅频响应示意图

7.2.2 滤波器组的设计

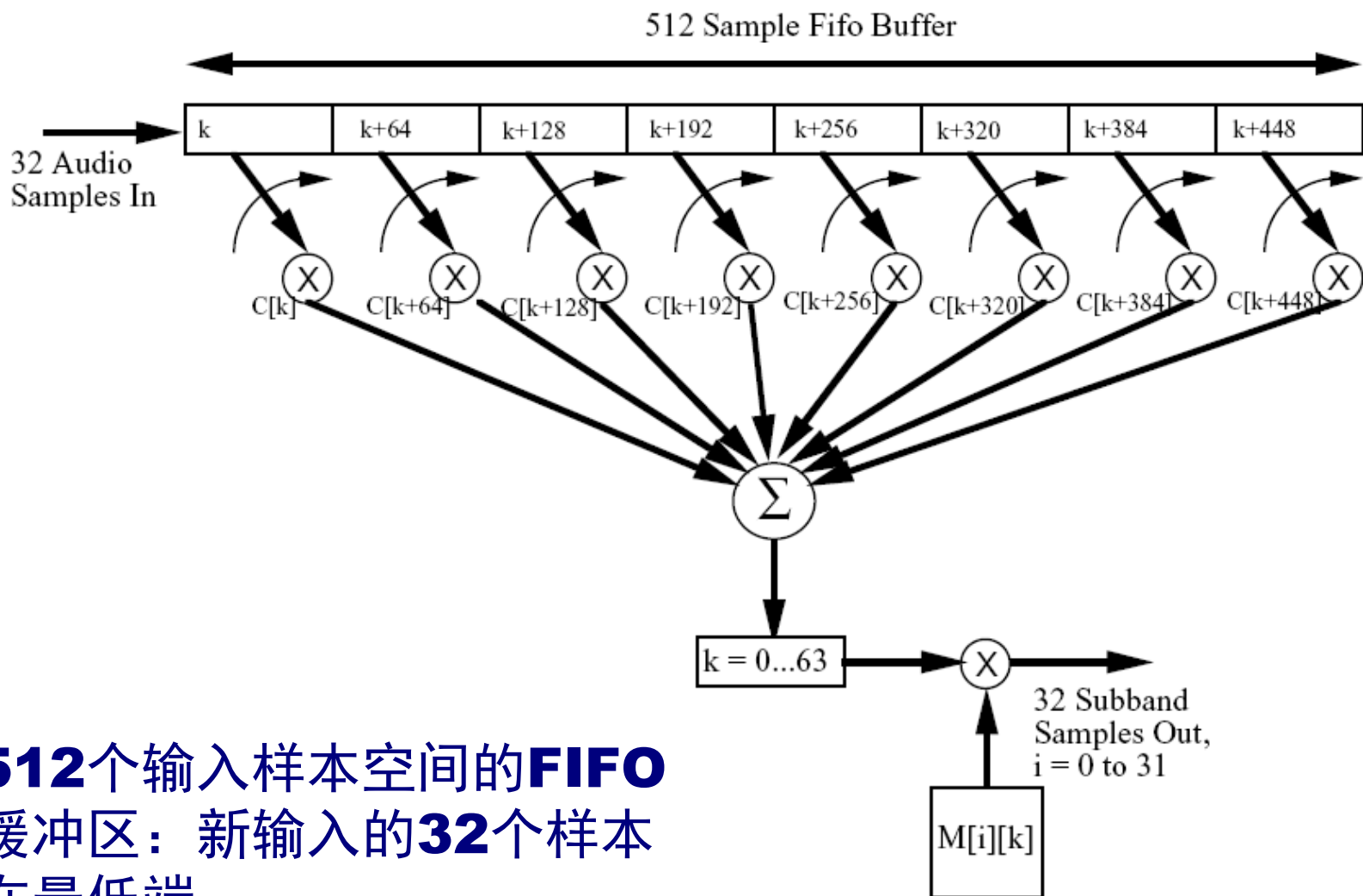
- 多相滤波器组(Polyphase Filter Bank): 将PCM样本变换到32个子带的频域信号
 - 由于cos项为周期对称函数, 表达式可合并cos项:

$$Y_{bp_n}(k) = \sum_{j=0}^7 \sum_{i=0}^{63} x(k - 2Mj - i)h(i + 2Mj)(-1)^j \cos[(2n - 1)i\pi/2M - \phi_n]$$

由于j与cos项无关, 进一步化简得

$$Y_{bp_n}(k) = \sum_{i=0}^{63} \left[\sum_{j=0}^7 x(k - 2Mj - i)h(i + 2Mj)(-1)^j \right] \cos[(2n - 1)i\pi/2M - \phi_n]$$

多相滤波器组



512个输入样本空间的**FIFO**缓冲区：新输入的**32**个样本在最低端

多相滤波器组(2)

- for $i=0\dots 511$, 计算窗口内的样本: $Z[i] = C[i] \cdot X[i]$
 - 标准中规定了分析窗口的512个系数 $C[i]$
- 样本点分组, 计算64个 Y_k 的值: $Y[k] = \sum_{j=0}^7 Z[k + 64j]$
- 计算32个子带样本: $S[i] = \sum_{k=0}^{63} Y[k] \cdot M[i][k]$
- 总计算公式:
$$S[i] = \sum_{k=0}^{63} \sum_{j=0}^7 M[i][k] (C[k + 64j] X[k + 64j])$$
 - 其中分析矩阵: $M[i][k] = \cos\left(\frac{(2i+1)(k-16)\pi}{64}\right)$
 - 共需: $512 + 32 \times 64 = 2560$ 次乘法
 - 每个子带的带宽为 $\pi/32T$, 中心为 $\pi/64T$ 的奇数倍

多相滤波器组(3)

- 多相滤波器组的公式

$$S[i] = \sum_{k=0}^{63} \sum_{j=0}^7 M[i][k] (C[k + 64j] X[k + 64j])$$

- 可用卷积公式代替：

$$S_i[i] = \sum_{n=0}^{511} X[n-t] h_i[n]$$

- 其中 $h_i[n] = h[n] \times \cos\left(\frac{(2i+1)(n-16)\pi}{64}\right)$

低通

调制

- 当 $n/64$ 的整数部分为奇数时，令 $h[n] = -C[n]$
- 反之， $h[n] = C[n]$
- 每个子带有自己的带通滤波器脉冲响应
 - 直观但效率不高：16384次乘法和16352次加法

多相滤波器组(4)

■ 缺点:

- 等带宽的滤波器组与人类听觉系统的临界频带不对应
 - 在低频区域，单个子带会覆盖多个临界频带。在这种情况下，量化比特数不能兼每个临界频带
- 滤波器组与其逆过程**不是**无失真的
 - 但滤波器组引入的误差很小，且听不到
- 子带间频率有混叠
 - 滤波后的相邻子带有频率混叠现象，一个子带中的信号可以影响相邻子带的输出

多相滤波器组(5)

- 混叠：一个单频正弦信号输入可能在两个子带中产生非零信号

Input audio: 1,500-Hz sine wave sampled at 32 kHz, 64 of 256 samples shown

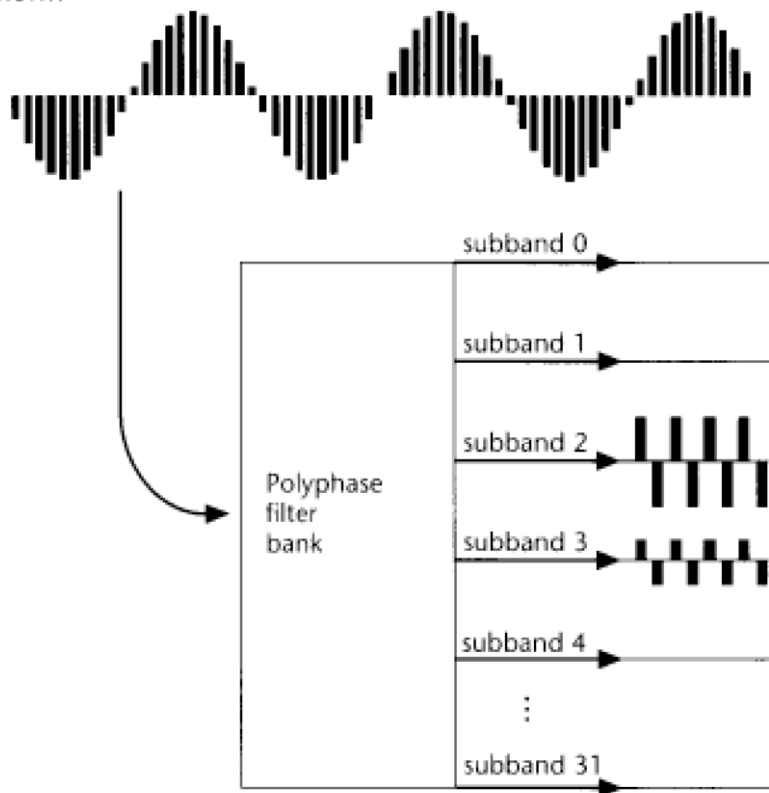
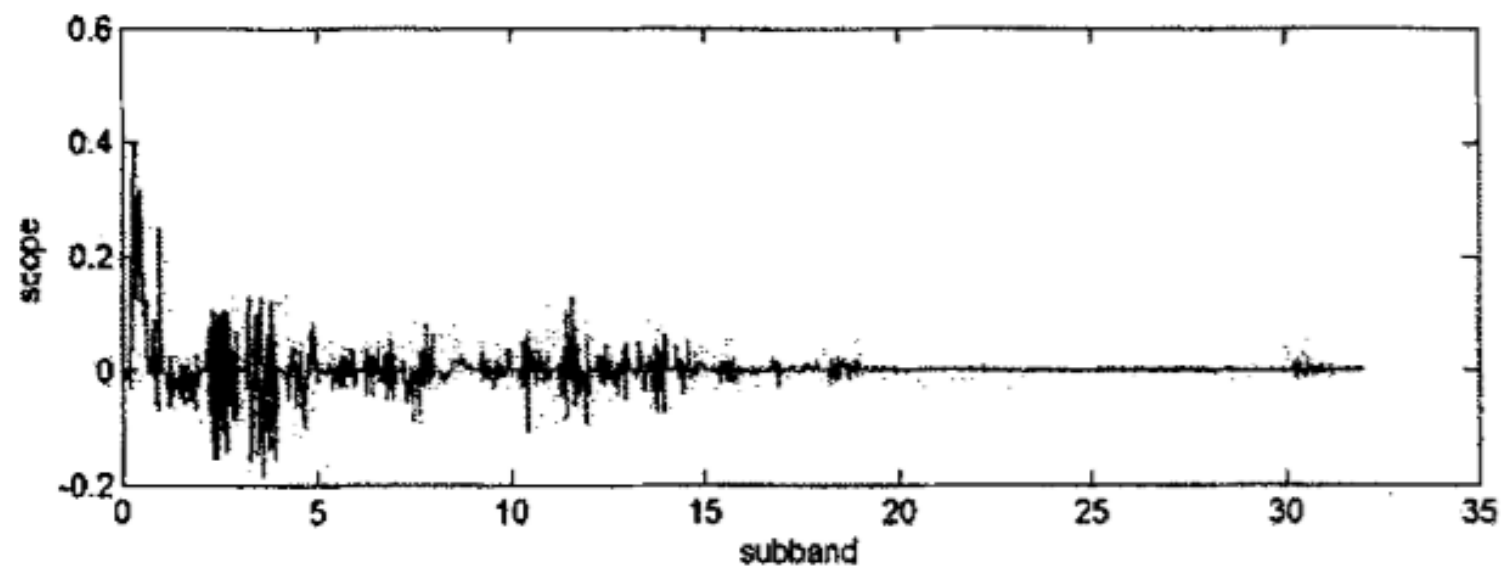
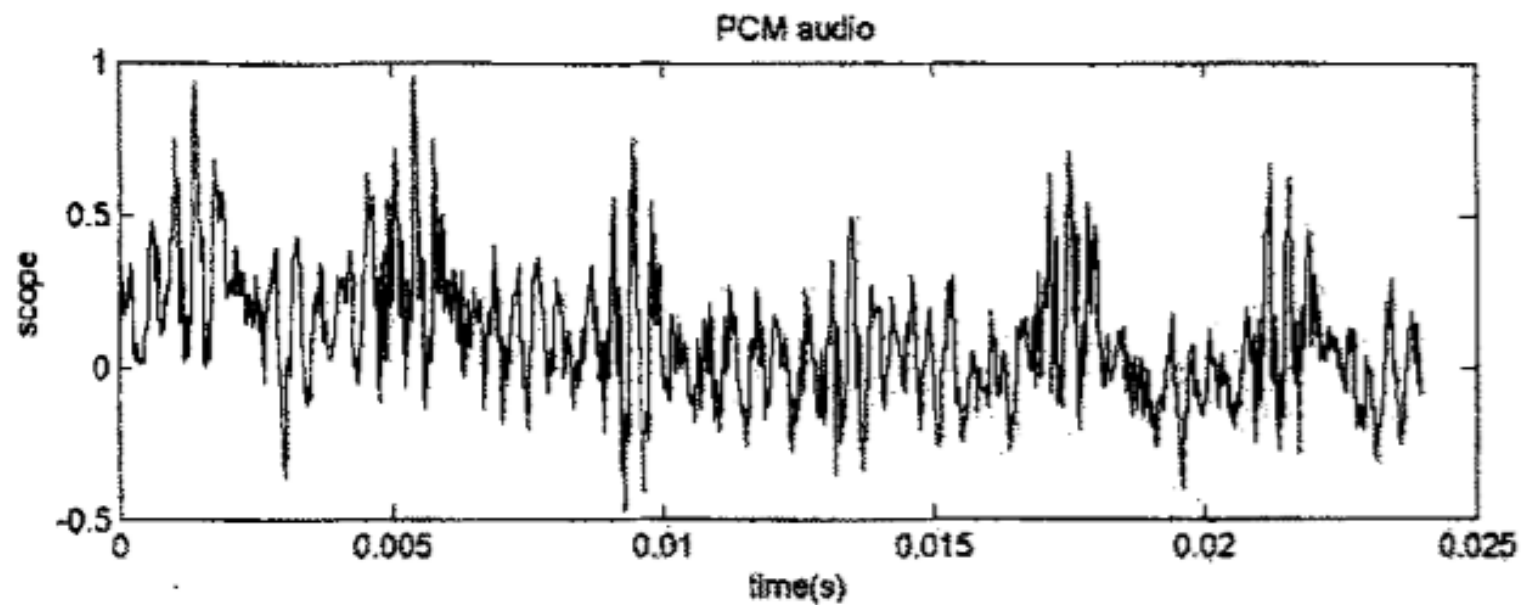


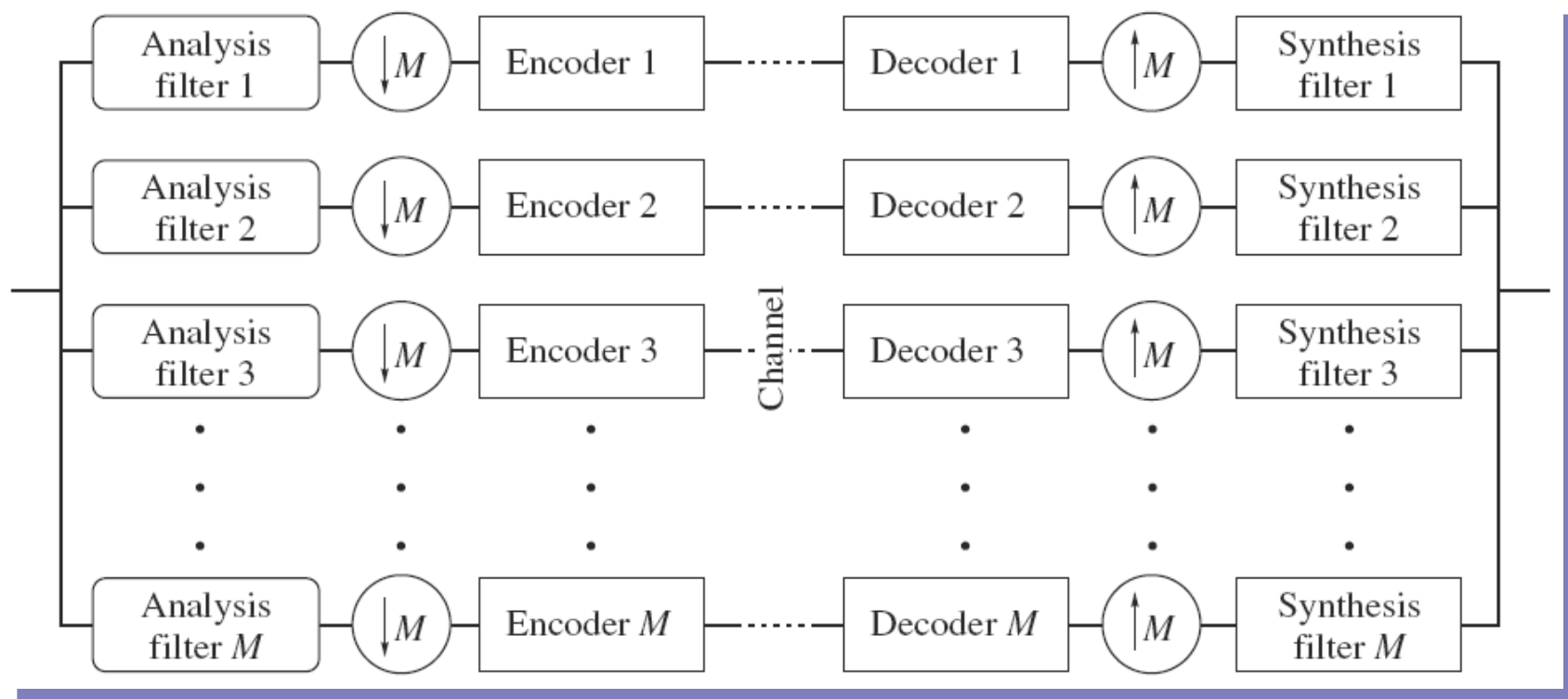
Figure 5. Aliasing: Pure sinusoid input can

Subband outputs:
8x32 samples; both subbands 3 and 4
have significant output values

多相滤波器组(6)



7.2.3 基本的子带编码系统



■ 三大要素：

- 分析—综合滤波器组
- 比特分配机制
- 编码机制：PCM、ADPCM、VQ...

7.2.3 基本的子带编码系统

■ 量化&编码

➤ 不同子带运载不同数量的感知信息

➤ 比特分配对总质量影响极大

➤ 感知边界之外的带宽可以被抛弃

■ 不太重要的子带可以用较低码率下采样，且可以采用较粗糙的量化粗量化

■ 运载最重要信息的子带用高质量编码

■ 计算比特分配 R_k ，使得给定码率 R 下失真 D 最小

$$R = \frac{1}{M} \sum_{k=1}^M R_k \quad J_k = D_k + \lambda_k R_k$$

语音编码： G.722

■ 目标

- 64 kbits/s的高质量编码
- 中等速率的编码系统，子带编码的动态范围宽、音质高、成本低
- 语音存储转发/语音邮件

■ 额外可选

- 48 kbps, 56 kbps (给辅助信道预留空间)

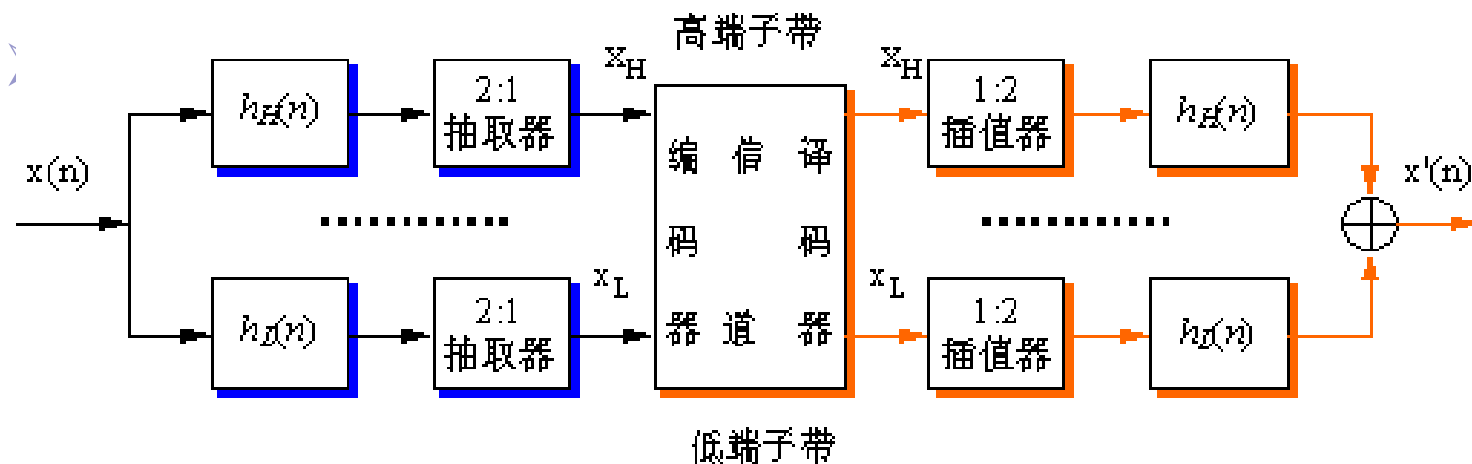
■ 输入信号为

- 以7KHz抗混叠滤波
 - 人的发音器官发出的声音频段在**80Hz到3400Hz**之间
- 采样率为16,000个样本/秒
- 用14比特的均匀量化器量化

语音编码： G.722 (2)

■ 输入信号

- 通过一组包含2个 24系数的FIR滤波器
 - 低通： 0-4kHz
 - 高通： 4kHz+
- 用2倍于子带带宽的采样频率对子带信号进行采样



语音编码： G.722 (3)

■ ADPCM

- 低频子带以6比特/样本编码
 - 可以选择去掉1-2个最不重要的比特
- 高频子带以2比特/样本编码

■ 量化器&预测器

- 基于Jayant机制的自适应量化，使用过去的2个重构值和6个量化器输出来预测下一个样本（同G.726）

■ 在接收方

- 反过程，通过插入0进行上采样

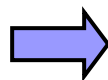
图像压缩中的应用

- 需要2-D滤波器
 - 通常通过可分解的滤波器实现
- 每行被滤波为低/高频部分
- 用2倍于子带带宽的采样频率采样
- 即：
 - 从 $N \times N$ 的图像开始，我们得到
 - $2 \times (N \times N/2)$ 图像
 - $4 \times (N/2 \times N/2)$ 图像
 - 可以进一步递归分解，但实际中很少用到

分解示例

对于第一个输入（也就是图像左边缘处的像素和上行的像素），假定过去的输入值为0

10	14	10	12	14	8	14	12
10	12	8	12	10	6	10	12
12	10	8	6	8	10	12	14
8	6	4	6	4	6	8	10
14	12	10	8	6	4	6	8
12	8	12	10	6	6	6	6
12	10	6	6	6	6	6	6
6	6	6	6	6	6	6	6



Decimated Low-Pass Output				Decimated High-Pass Output			
5	12	13	11	5	-2	1	3
5	10	11	8	5	-2	-1	2
6	9	7	11	6	-1	1	1
4	5	5	7	4	-1	-1	1
7	11	7	5	7	-1	-1	1
6	10	8	6	6	2	-2	0
6	8	6	6	6	-2	0	0
3	6	6	6	3	0	0	0

Low-Low Image				Low-High Image			
2.5	6	6.5	5.5	2.5	6	6.5	5.5
5.5	9.5	9	9.5	0.5	-0.5	-2	1.5
5.5	8	6	6	1.5	3	1	-1
6	9	7	6	0	-1	-1	0

High-Low Image				High-High Image			
2.5	-1	0.5	1.5	2.5	-1	0.5	1.5
5.5	-1.5	0	1.5	0.5	0.5	1	-0.5
5.5	-1	-1	1	1.5	0	0	0
6	0	-1	0	0	-2	1	0

注意：
滤波操作中过去值的影响可颠倒边界处的像素值，作为假设的过去值

对于序列6, 9, 5, 4, 7, 2, ...。使用一个三抽头滤波器对其进行滤波，则增加两个过去值9 6 6 9 5 4 7 2

例7.5 分解与合成滤波器例

(5, 3) filter:

$$h_0[n] = (-1 \ 2 \ 6 \ 2 \ -1)/8 \quad g_0[n] = (1 \ 2 \ 1)/2$$

$$h_1[n] = (-1 \ 2 \ -1)/2 \quad g_1[n] = (-1 \ -2 \ 6 \ -2 \ -1)/8$$

input: 100, 100, 100, 100, 200, 200, 200, 200, ...

先向左扩展两个样点:

$$(-100+200+600+400-200)/8$$

100, 100, 100, 100, 100, 100, 200, 200, 200, 200, ...

低通: 100 100 87.5 112.5 187.5 212.5 200 200

高通: 0 0 0 -50 50 0 0 0

还原:

低通: 100 100 106.25 112.5 162.5 212.5 206.25 200

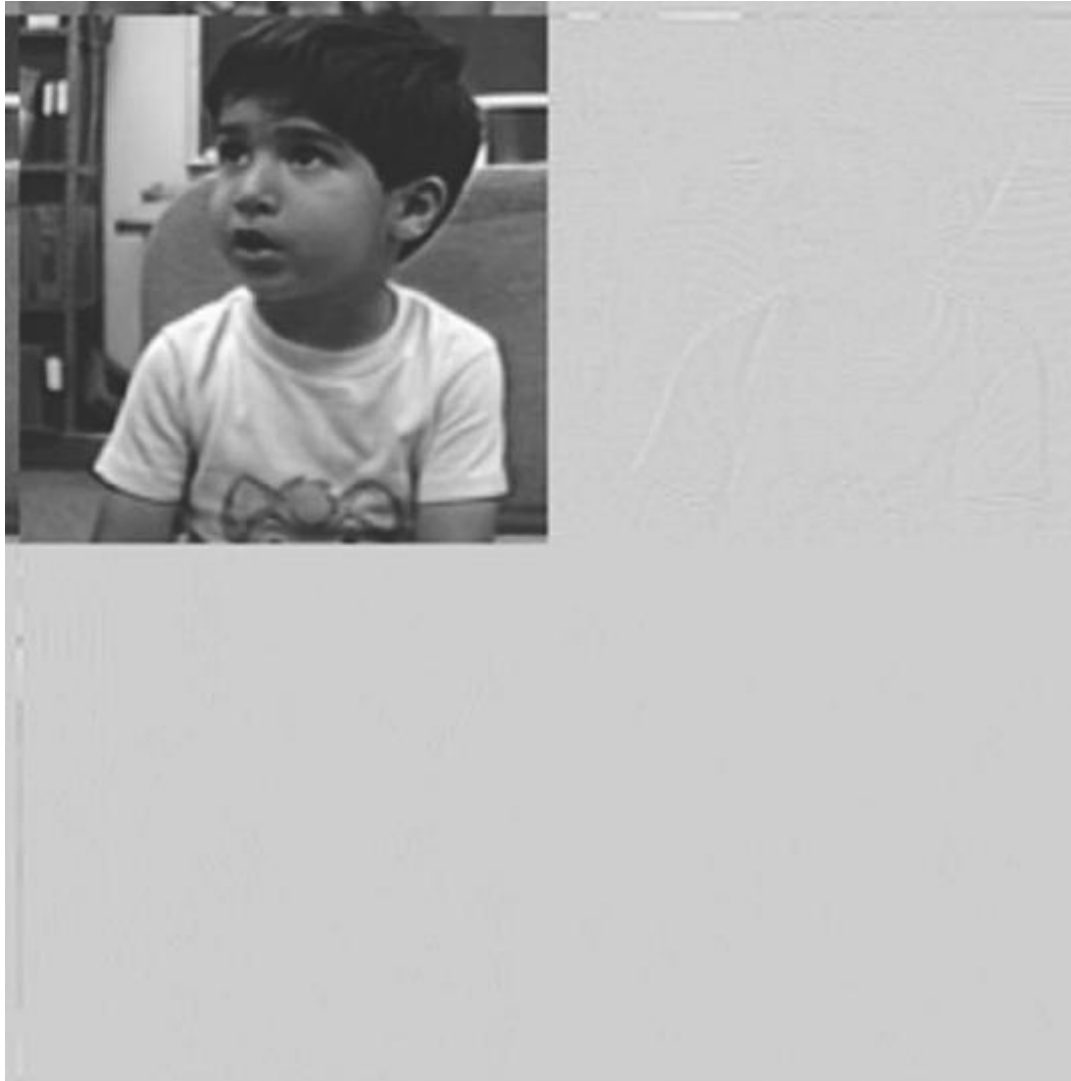
高通: 0 0 -6.25 -12.5 37.5 -12.5 -6.25 0

相加: 100 100 100 100 200 200 200 200

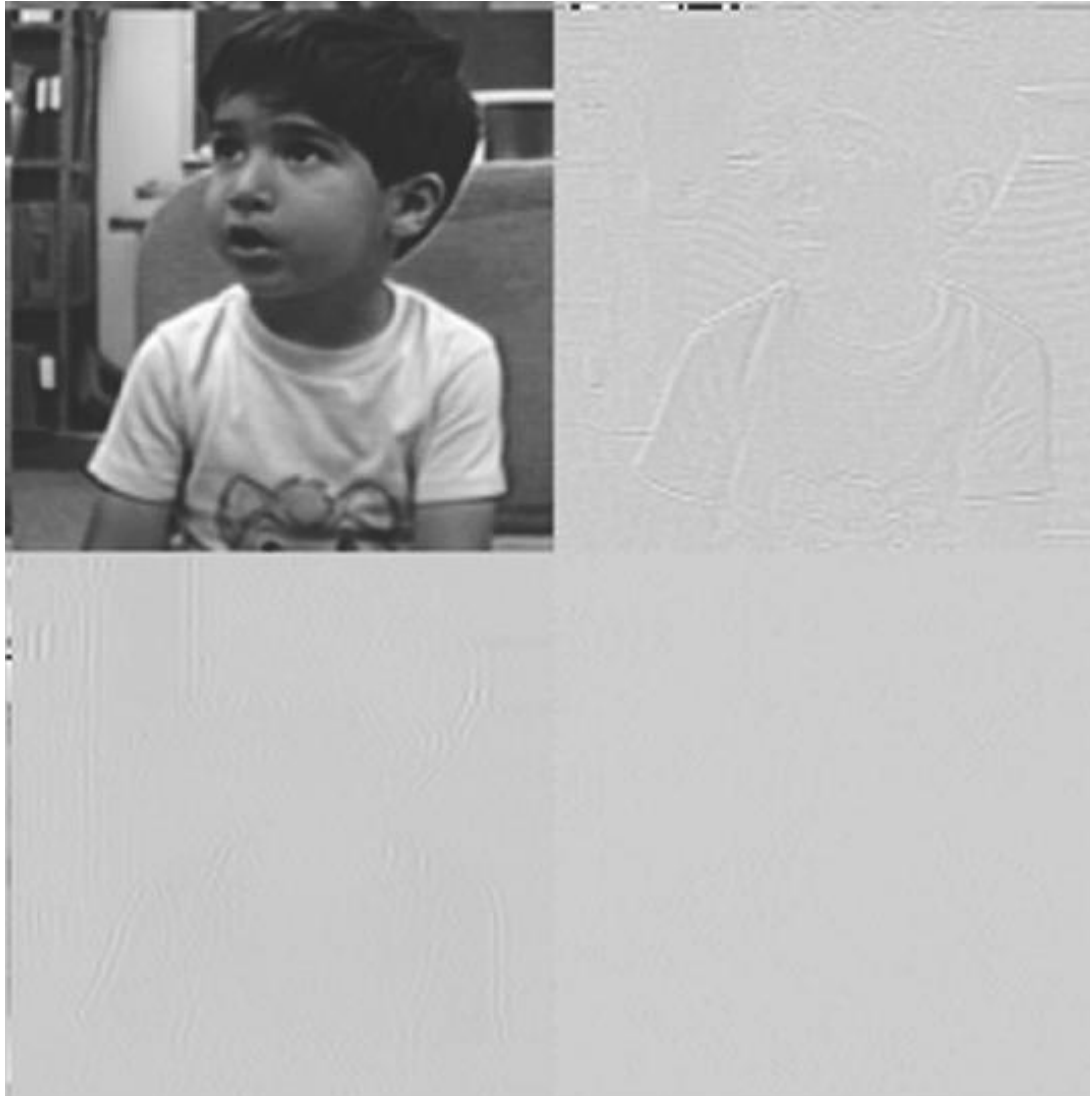
图像分解：8-tap Johnston



图像分解： 16-tap Johnston



图像分解：8-tap Smith/Barnwell



压缩对重构的影响



0.5 bpp, 8-tap Johnston



0.5 bpp, 8-tap Smith/Barnwell

为不同子带递推分配比特。在使用johnson滤波器分解时，高低子带中存在大量能量，需要各分配1bit

在使用Smith/Barnwell滤波器时，除低低子带之外的所有子带中都没有什么内容，所以分配了2bit

总结

■ 子带编码:

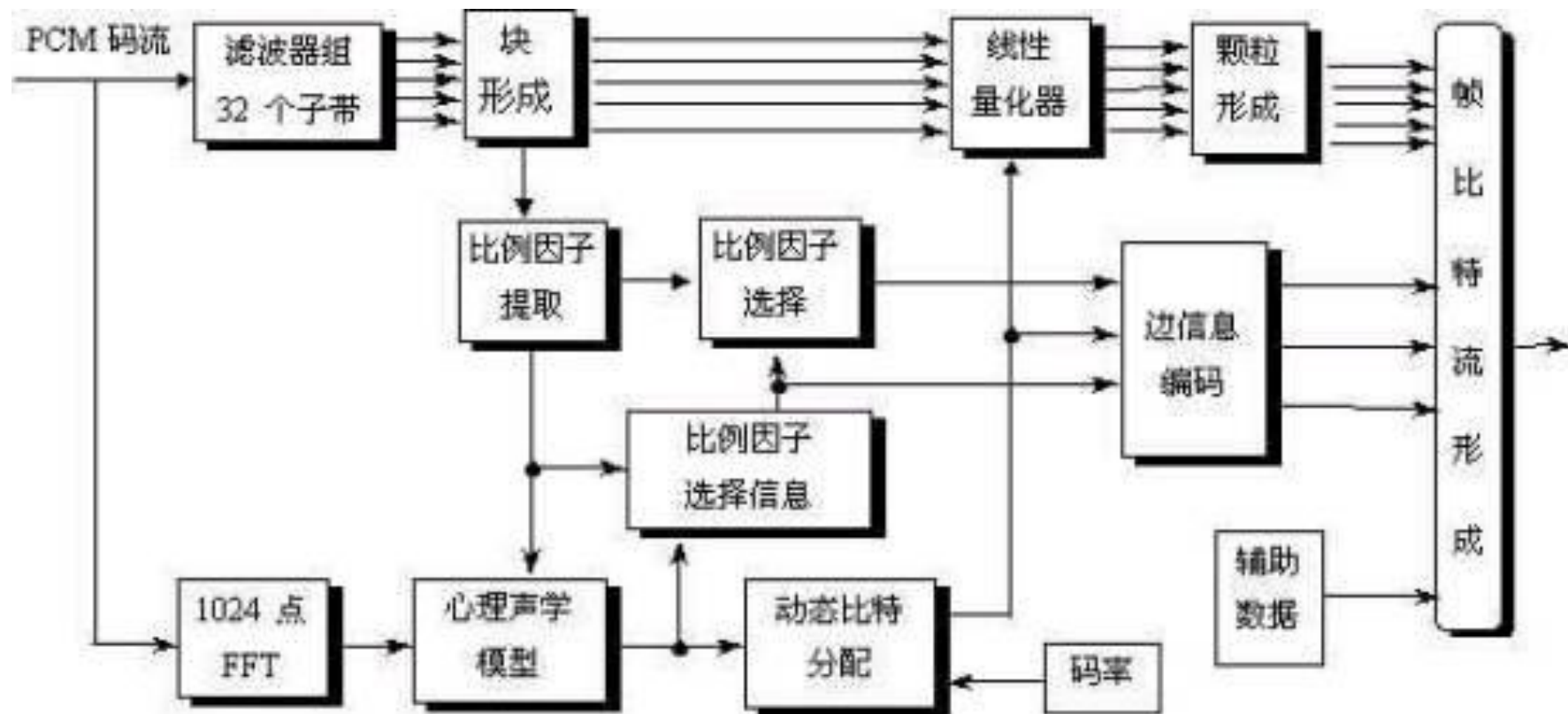
- 将输入信号分解成子部分
- 对每个子部分用合适的方式编码

■ 编码过程:

- 选择分解滤波器
- 利用选定滤波器，计算子带输出
- 对滤波器输出进行下采样
- 对采样后的输出编码

■ 解码为编码的逆过程

7.2.4 感知音频编码原理



感知编码的基本思路

声音压缩算法可以确立这种特性的模型来取消更多的冗余数据

■ 基本思想

- 分析信号，去掉不能被感知的部分

■ 心理声学模型(Pschoacoustic model)

- 生理(Physiological)感知极限（传感极限）
- 心理 (Psychological)感知极限（信号处理极限）

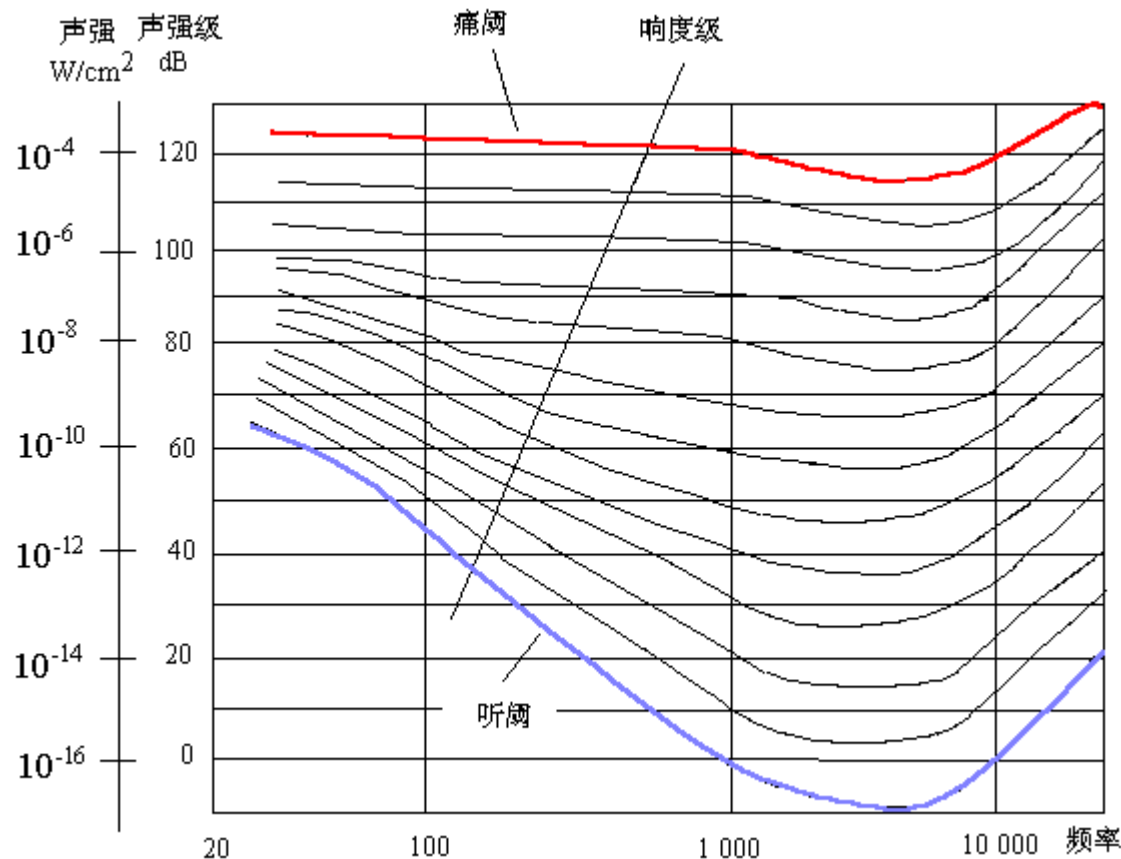
■ 心理声学模型

- 听觉系统中存在一个听觉阈值电平，低于这个电平的声音信号就听不到
 - 听觉阈值的大小随声音频率的改变而改变。
 - 一个人是否听到声音取决于声音的频率，以及声音的幅度是否高于这种频率下的听觉阈值
- 听觉掩蔽特性。即听觉阈值电平是自适应的，会随听到的不同频率声音而发生变化
 - 一种是1000Hz的声音，另一种是1100Hz的声音。但它的强度比前者低18dB？

MPEG-I 心理声学模型

- 通过子带分析滤波器组使信号具有高的时间分辨率，确保在短暂冲击信号情况下，编码的声音信号具有足够高的质量
- 又可以使信号通过FFT运算具有高的频率分辨率，因为掩蔽阈值是从功率谱密度推出来的。
- 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

听觉阈值



- 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。

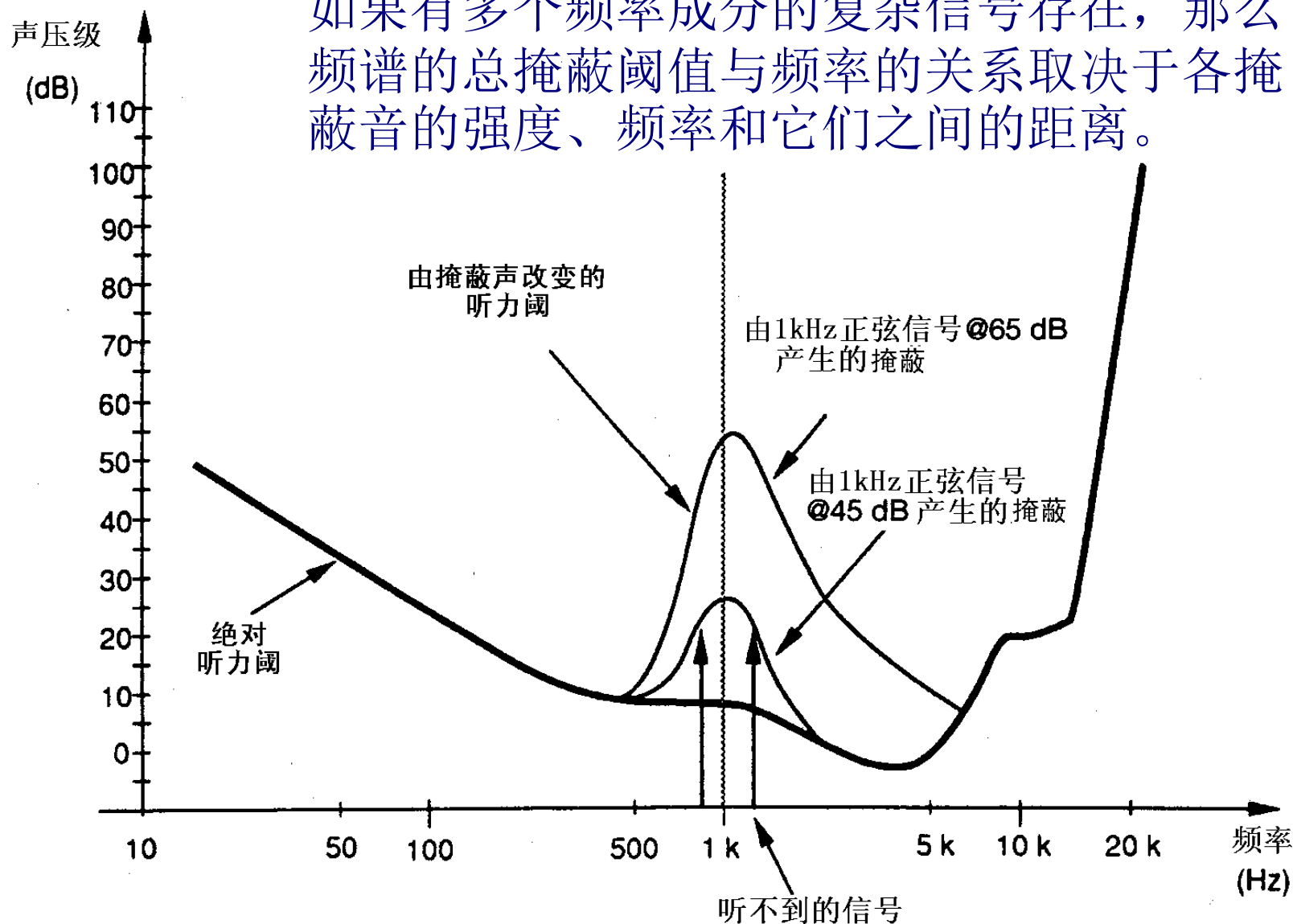
■ 声压级(SPL):

$$SPL = 20 \lg \frac{P_{rms}}{P_{ref}} \quad (\text{分贝dB})$$

➤ $P_{ref} = 2 \times 10^{-5}$ ，零声级的参考声压值，为具有正常听力的年轻人对1 kHz的声音刚好能察觉的声压值

频域掩蔽域随声压级变化曲线

如果有多个频率成分的复杂信号存在，那么频谱的总掩蔽阈值与频率的关系取决于各掩蔽音的强度、频率和它们之间的距离。



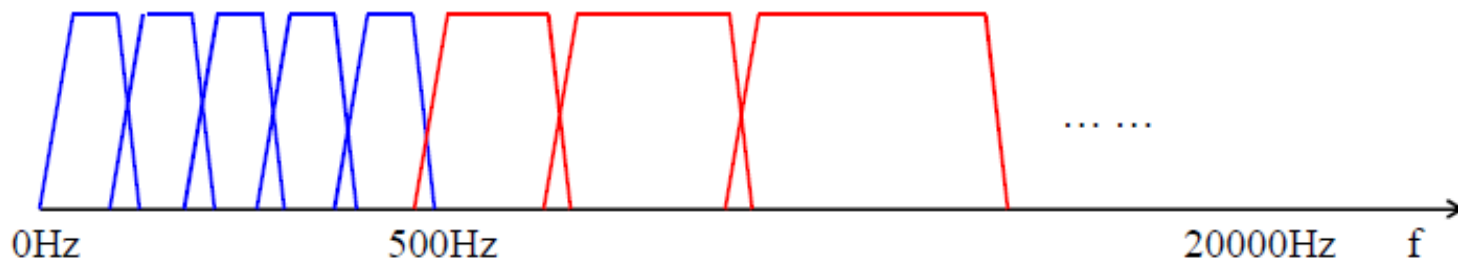
临界频带(Critical Band)

- 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。
- 研究窄带噪声对纯音掩蔽量的规律时发现的
 - 使噪声的中心频率等于信号频率，只改变噪声的带宽同时保持噪声的功率谱密度不变，测试纯音听阈随掩蔽噪声带宽变化的特性
 - 纯音的听阈随掩蔽噪声带宽的增大而增大，在带宽增大到某一特定值之后听阈保持恒定不变。
- 通常认为从20Hz到16kHz有25个临界频带，单位为bark，1 Bark = 一个临界频带的宽度

$$1 \text{ Bark} = \begin{cases} \frac{f}{100}, & \text{for frequencies } f < 500 \text{ Hz,} \\ 9 + 4 \log \left(\frac{f}{1000} \right), & \text{for frequencies } f \geq 500 \text{ Hz.} \end{cases}$$

人耳听觉系统

- 人类听觉系统大致等效于一个信号通过一组并联的不同中心频率的带通滤波器
 - 中心频率与信号频率相同的滤波器具有最大响应；中心频率偏离信号频率较多的滤波器不会产生响应。
 - 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。
 - 听音者在噪声中听某一纯音信号时，只启用中心频率与信号频率相同的那个听觉滤波器，纯音信号通过该滤波器，而噪声信号只有通带范围内的部分信号能通过，通带以外的频率成分则被抑制，只有通过该滤波器的噪声才对掩蔽起作用。
 - 聆听复音时启动多个听觉滤波器。听觉能够计算各滤波器输出端的信噪比。当信噪比达到或者超过听阈因子时，即可听到该频率成分。



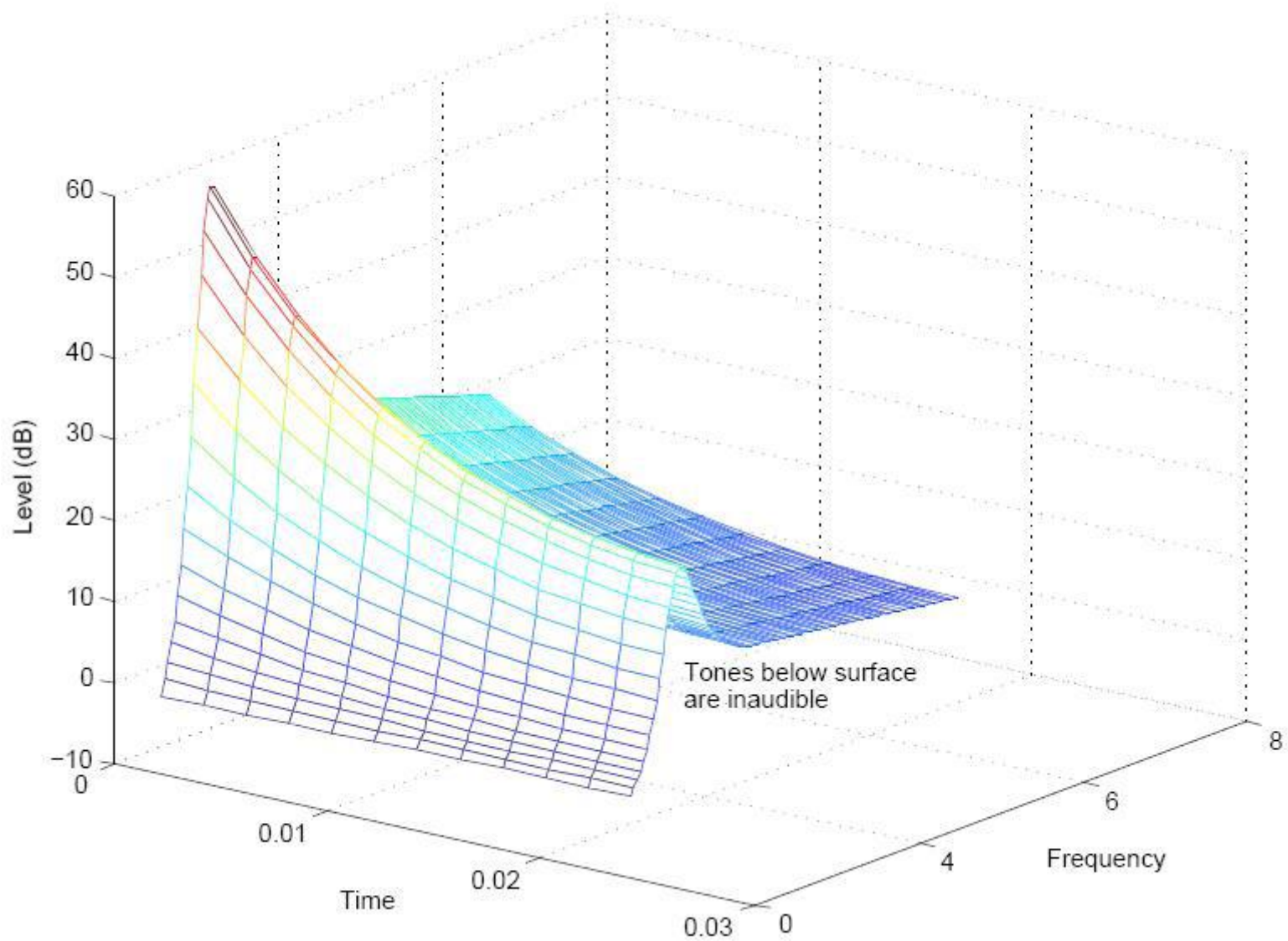
人耳听觉系统

- 掩蔽声对被掩蔽声的掩蔽效应，取决于两者的频率与强度的关系
 - 噪声掩蔽纯音时，只有以纯音频率为中心的，一定频带宽度内的噪声能量起掩蔽作用，超出该频带的噪声能量无掩蔽效应，称掩蔽的临界带宽。掩蔽说明了频率选择性的极限。
 - 掩蔽纯音信号，理论上使用带宽范围等于其临界带宽1/3倍频程的窄带噪声，这是因为掩蔽的临界带宽稍窄，听起来接近于纯音，患者常常混淆纯音与掩蔽噪声，而以纯音频率为中心频率的1/3倍频程噪声，宽度略大于临界带宽，同样可以起到很好的掩蔽效果。但实际应用中根据IEC645规定，临床听力计上常使用约4/10倍频程的窄带噪声。
- 掩蔽言语声和短声等宽频谱信号，则常用白噪声。因此，在听觉诱发反应测试中，往往在健侧耳施加白噪声作为掩蔽噪声。

临界频带(Critical Band)

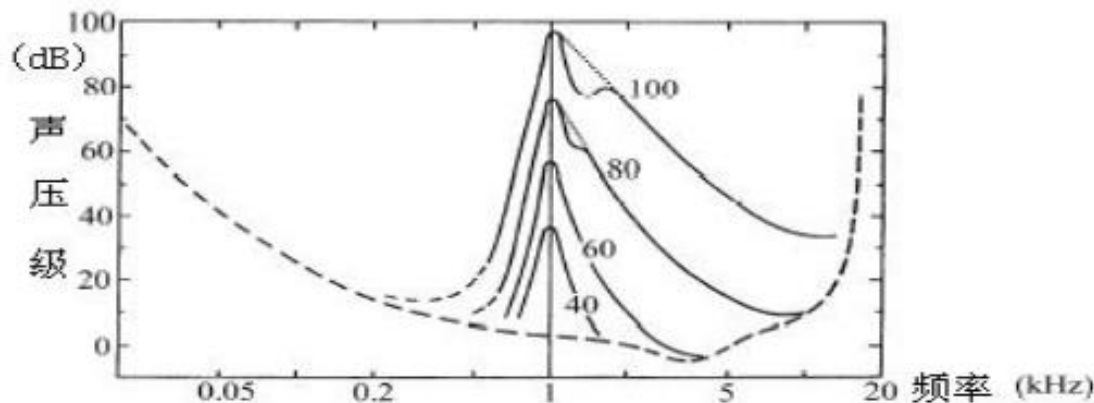
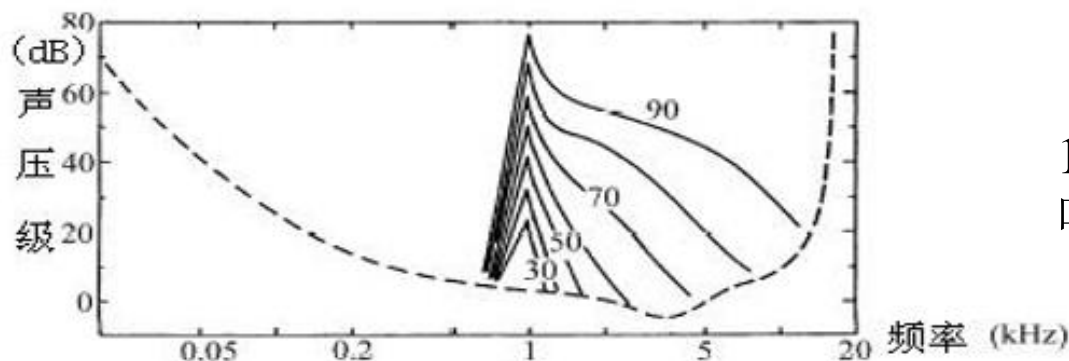
临界	频率 (Hz)			临界	频率 (Hz)		
频带	低端	高端	宽度	频带	低端	高端	宽度
0	0	100	100	13	2000	2320	320
1	100	200	100	14	2320	2700	380
2	200	300	100	15	2700	3150	450
3	300	400	100	16	3150	3700	550
4	400	510	110	17	3700	4400	700
5	510	630	120	18	4400	5300	900
6	630	770	140	19	5300	6400	1100
7	770	920	150	20	6400	7700	1300
8	920	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				

时域/频域掩蔽



频域不同掩蔽类型的讨论

- 音频信号通常有较为复杂的频谱结构 因此能产生掩蔽阈值的掩蔽音分量也有许多。
- 掩蔽音与被掩蔽音的组合方式有四种，即它们分别可以是乐音信号（弦信号，**Tone**）或窄带噪声（**noise**）。乐音信号和窄带噪声信号作为掩蔽音时产生的掩蔽效果有很大不同



掩蔽效果的加和

- 音乐与语音信号大都由一系列复杂的频谱分量构成，相应的这些多个掩蔽分量也会相互影响并最终获得一个整体的掩蔽阈值。对于多个掩蔽分量的综合掩蔽效果，目前有多种模型给出了描述方案。
- Lutfi 对多个掩蔽音同时存在时的综合掩蔽效果进行了研究：每个掩蔽音的掩蔽效果先独立变换然后再线性相加。

$$F(M_{AB}) = F(M_A) + F(M_B)$$

$$F(M_A) = (10^{M_A/10})^p \quad p \text{取} 0.2-0.5$$

- 问题：综合掩蔽效果与各掩蔽信号的声音级以及掩蔽信号的个数都有关系。如何整合掩蔽音？
 - 当两个信号重叠并落在一个临界频带中时，二者的掩蔽分量可以线性相加。
 - 对于复杂音频信号可将其频谱分割成一系列离散段，每段就是一个掩蔽信号。各掩蔽音互不重叠，即以一个临界带为单位。各掩蔽音的声压级则通过将对应的临界频带上的短时功率谱密度线性相加得到。

掩蔽与量化

- 将输入信号变换到频域，再将结果分解成一些尽量与临界频带尽可能相似的子带，然后对每个子带进行量化，量化方式应当使得量化噪声听不见
- 例：分解成32个子带（前16个子带），各子带的能量表：

Band	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Level	0	8	12	10	6	2	10	60	35	20	15	2	3	5	3	1

- 60dB的第8子带的信号会对第7子带带去12 dB的掩蔽，对第9子带带去15dB (根据心理声学模型)
 - 第7子带的水平为10 dB (< 12 dB)，因此忽略
 - 第9子带的水平为35 dB (> 15 dB)，因此需要发送
 - 只需发送掩蔽水平以上的能量。因此无需6 比特编码，而只用4比特就够——节省2个比特 (= 12 dB)
 - “用于表示系数的比特数使得由于量化引起的噪声低于掩蔽效应” [量化引起噪声 = 12dB; 掩蔽 = 15 dB]

7.2.5 MPEG音频压缩

1. MPEG-1声音的主要性能

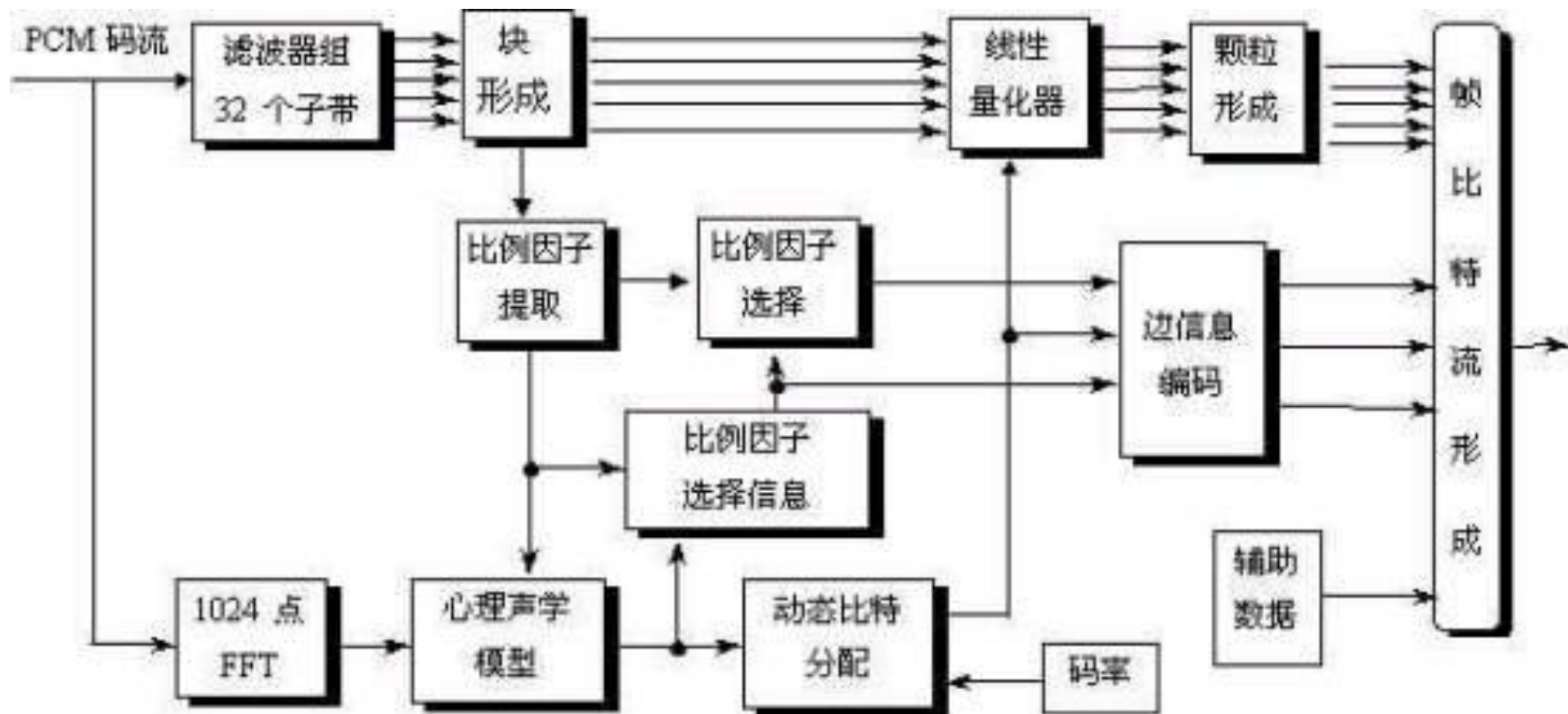
- 输入为PCM信号，采样率为32，44.1或48kHz，输出为32kbps到384kbps。
- 三个独立的压缩层次
 - Layer1:** 编码器最简单，384kbps（4: 1，用于小型数字盒带DCC，Compact Cassette）
 - Layer2:** 编码器复杂程度中等，256kbps~192kbps（6: 1~8: 1，用于DAB、CD-I和VCD）
 - Layer3:** 编码器最为复杂，64kbps，用于ISDN，网络音频。

7.2.5 MPEG音频压缩

MPEG1音频压缩的数据率

音质要求	声音带宽 (kHz)	方式	数据率 (kbps)	压缩比
电话	2.5	单声道	8	96 : 1
优于短波	5.5	单声道	16	48 : 1
优于调幅广播	7.5	单声道	32	24 : 1
类似于调频广播	11	立体声	56 ~ 64	26 ~ 24 : 1
接近CD	15	立体声	96	16 : 1
CD	> 15	立体声	112 ~ 128	12 ~ 10 : 1

7.2.5 MPEG音频压缩



7.2.5 MPEG音频压缩

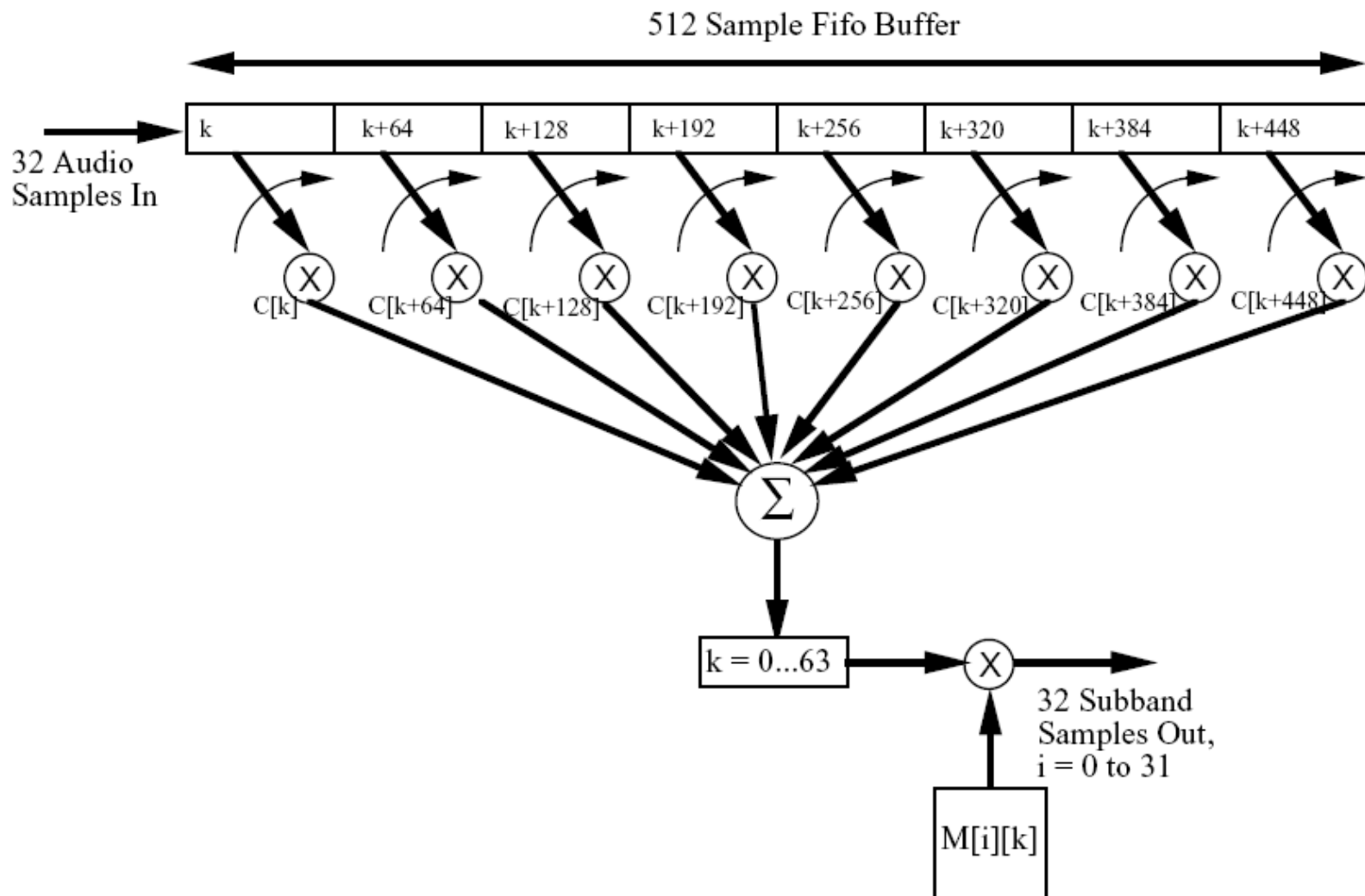
2. 编码器说明

输入声音信号经过一个多相滤波器组，变换到多个子带。同时经过“心理声学模型”计算以频率为自变量的噪声掩蔽阈值。量化和编码部分用信掩比**SMR**决定分配给子带信号的量化位数，使量化噪声 $<$ 掩蔽域值。最后通过数据帧包装将量化的子带样本和其它数据按照规定的帧格式组装成比特数据流。

- (1) 多相滤波器组，用来分割子带
- 划分子带的方法有两种：线性划分和非线性划分
 - 线性划分可能一个子带覆盖好几个临界频带

(1) 多相滤波器组，用来分割子带

以Layer 1为例，先分成32个相等的子带。



(2) 量化和编码 – 比例因子的取值和编码

对各个子带每12个样点进行一次比例因子计算。先定出12个样点中绝对值的最大值。查比例因子表中比这个最大值大的最小值作为比例因子。用6比特表示。

◆ 第2层的一帧对应36个子带样值，是第1层的三倍，原则上要传三个比例因子。为了降低比例因子的传输码率，采用了利用人耳时域掩蔽特性的编码策略。

◆ 每帧中每个子带的三个比例因子被一起考虑，划分成特定的几种模式。根据这些模式，1个、2个或3个比例因子和比例因子选择信息（每子带2比特）一起被传送。如果一个比例因子和下一个只有很小的差别，就只传送大的一个，这种情况对于稳态信号经常出现。

◆ 使用这一算法后，和第1层相比，第2层传输的比例因子平均减少了2个，即传输码率由22.5Kb/s降低到了7.5Kb/s。

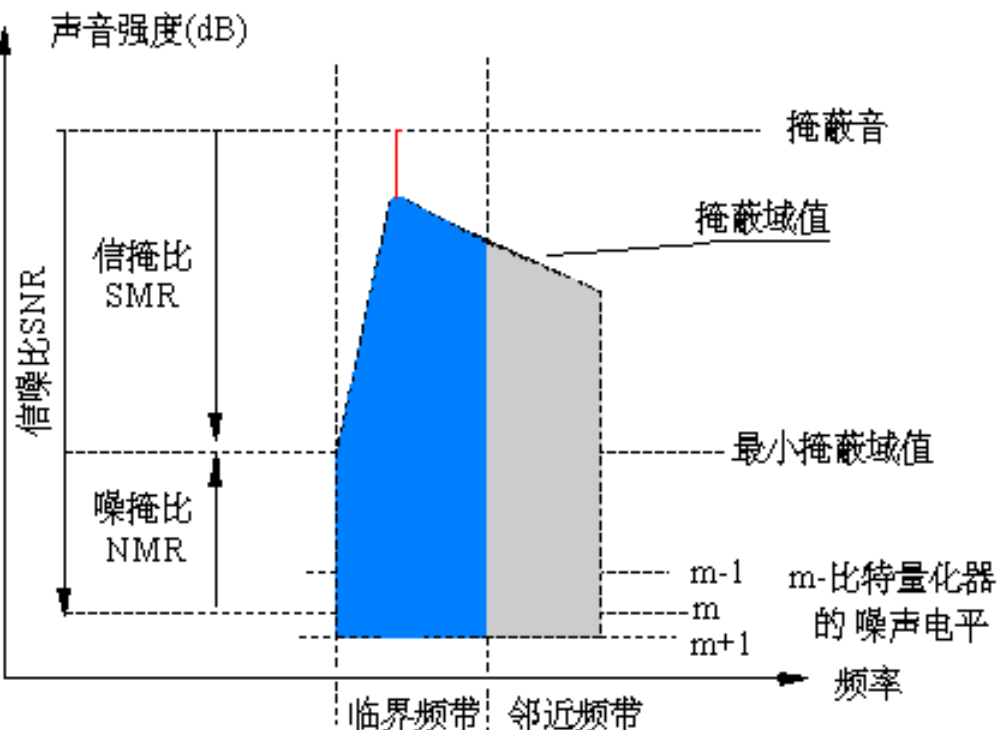
(2)量化和编码 – 比特分配及编码

■ 在调整到固定的码率之前

- 先确定可用于样值编码的有效比特数
- 这个数值取决于比例因子、比例因子选择信息、比特分配信息以及辅助数据所需比特数

■ 比特分配的过程

- 对每个子带计算掩蔽-噪声比MNR，信噪比SNR-信掩比SMR，即： $MNR = SNR - SMR$



(2)量化和编码 – 比特分配及编码

■ 比特分配的过程

- 使整个一帧和每个子带的总噪声-掩蔽比最小。这是一个循环过程，每一次循环使获益最大的子带的量化级别增加一级，当然所用比特数不能超过一帧所能提供的最大数目
- 第1层一帧用4比特给每个子带的比特分配信息编码；而第2层只在低频段用4比特，高频段则用2比特

(2)量化和编码 – 子带样值的量化和编码

- 输入以12个样本为一组，每组样本经过时间-频率变换之后进行一次比特分配并记录一个比例因子(scale factor)
- 比特分配信息告诉解码器每个样本由几位表示，比例因子用6比特表示，解码器使用这个6比特的比例因子乘逆量化器的每个输出样本值，以恢复被量化的子带值。比例因子的作用是充分利用量化器的量化范围，通过比特分配和比例因子相配合，可以表示动态范围超过120dB的样本。
- 第2层中，量化级别的数目随子带的不同而不同，但量化等级仍然覆盖了3~65535的范围，同时子带不被分配给比特的概率增加了，没有分配给比特的子带就不被量化。低频段的量化等级有15级，中频段7级，高频段只有3级

(3) 数据帧包装

■ 码流结构的设计

- 有利于使用低复杂性和少延时的解码器，编码后的声音信号包含多个短且恒定间隔的切入点
- 编码数据允许在编码码流中插入整数倍的切入点，以利记录、播放和编辑短的声音序列，并能精确指定编辑点

(3) 数据帧包装

■为了能简单地在解码器中实现上述功能，切入点的帧必须包括码流解码所需的完整信息

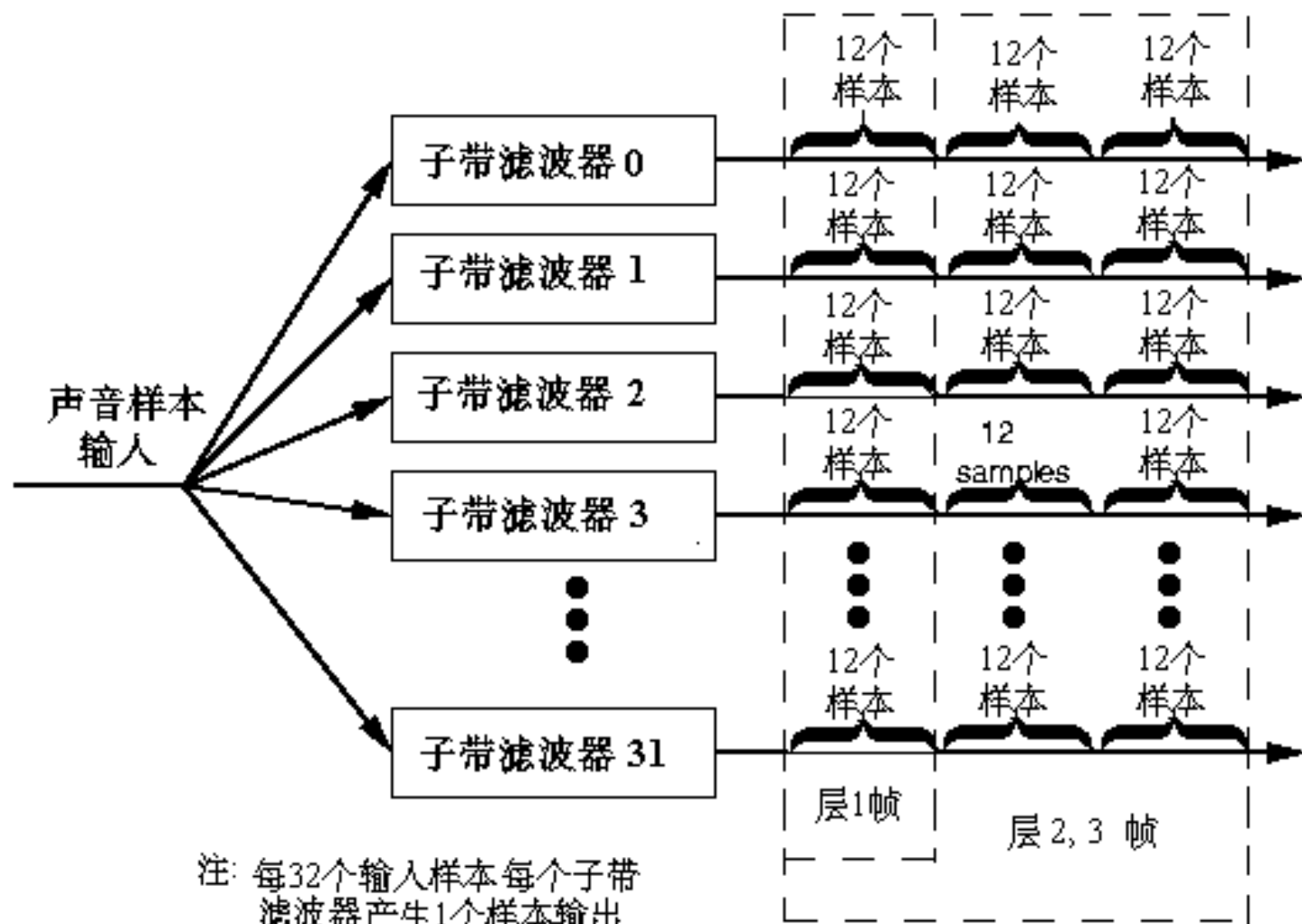
■这些特性在数字声音广播领域也很重要，因为由于经济原因须要有低复杂性的解码器，而且码流中要有许多切入点，以易于对传输误码造成的连续样值错误进行块隐藏。

帧头	CR C	比特 分配	SCFSI	比例 因子	样值	辅助数 据
----	---------	----------	-------	----------	----	----------

层II

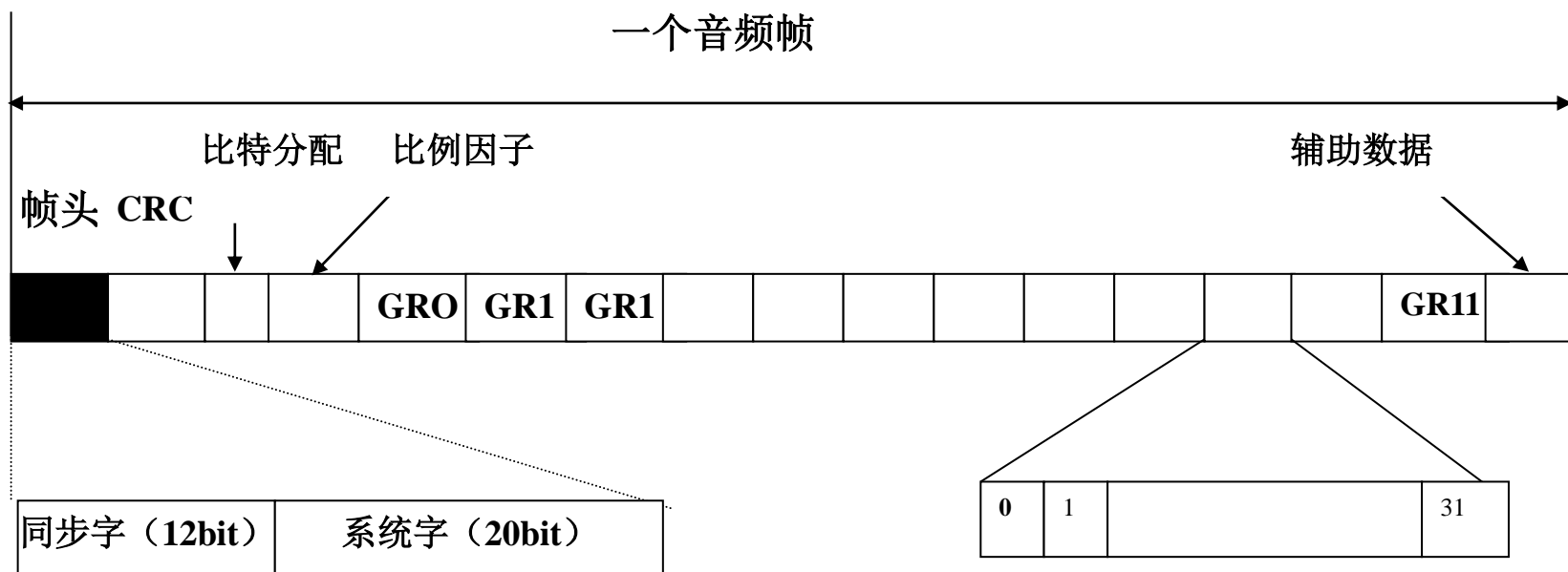
(32) (0,16) (26-188) (0-60) (0-1080) (1152)

3) 数据帧包装



3) 数据帧包装

- 层I每帧含384个样本数据。每帧由32个子带分别输出的12个样本值组成
 - 以48kHz采样，一帧长为 $32 \times 12 \times 20.83\mu s = 8ms$ 。
- 层II每帧包含1152个样本。低、中、高频段对比特分配不同，分别用4、3、2比特。比特流中增加了一个比特因子选择信息域，解码器根据这个域的信息可知道是否需要以及如何共享比例因子。



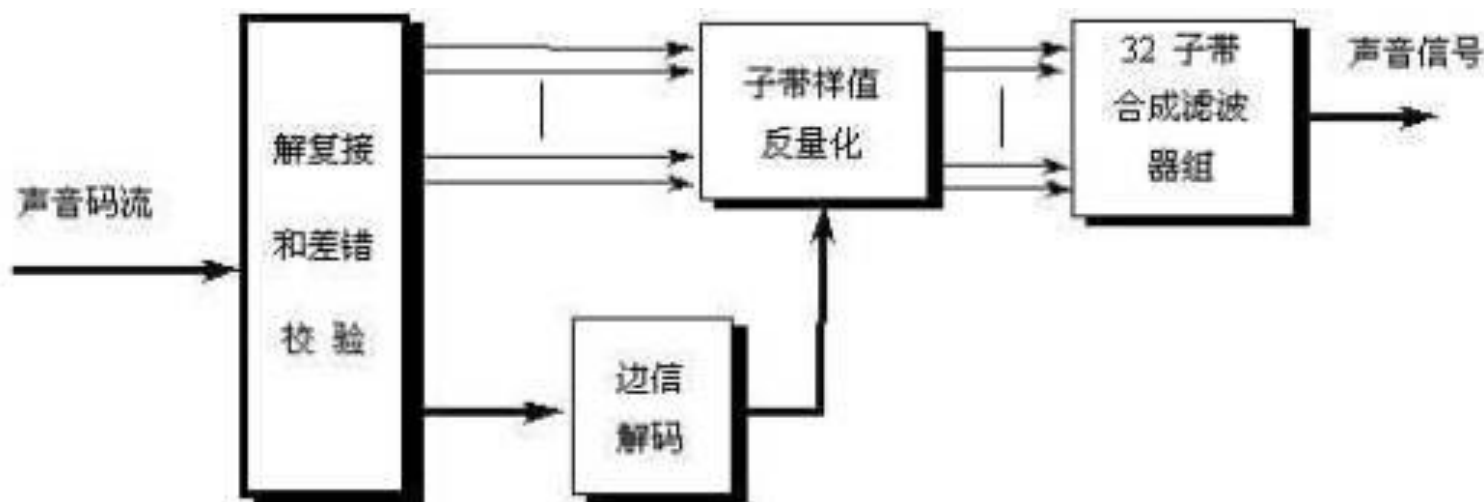
(3) 数据帧包装

帧头	CRC	比特分配	SCFSI	比例因子	样值	辅助数据	层II
(32)	(0,16)	(26-188)	(0-60)	(0-1080)	(1152)		

- 帧头(Header): 每帧开始的头32个比特, 包含有同步和状态比特流信息, 在所有层都相同,同步码字为12bit全1码(1111, 1111, 1111)。
- 误码检测CRC: 使用一种16bit奇偶校验字, 可供在比特流中作检测误码用。在所有层都相同
- 声音数据: 由比特分配表、比例因子选择信息、比例因子和子带样值组成, 其中子带样值是声音数据的最大部分。每层声音数据不同。
- 辅助数据: 用作辅助数据比特流。

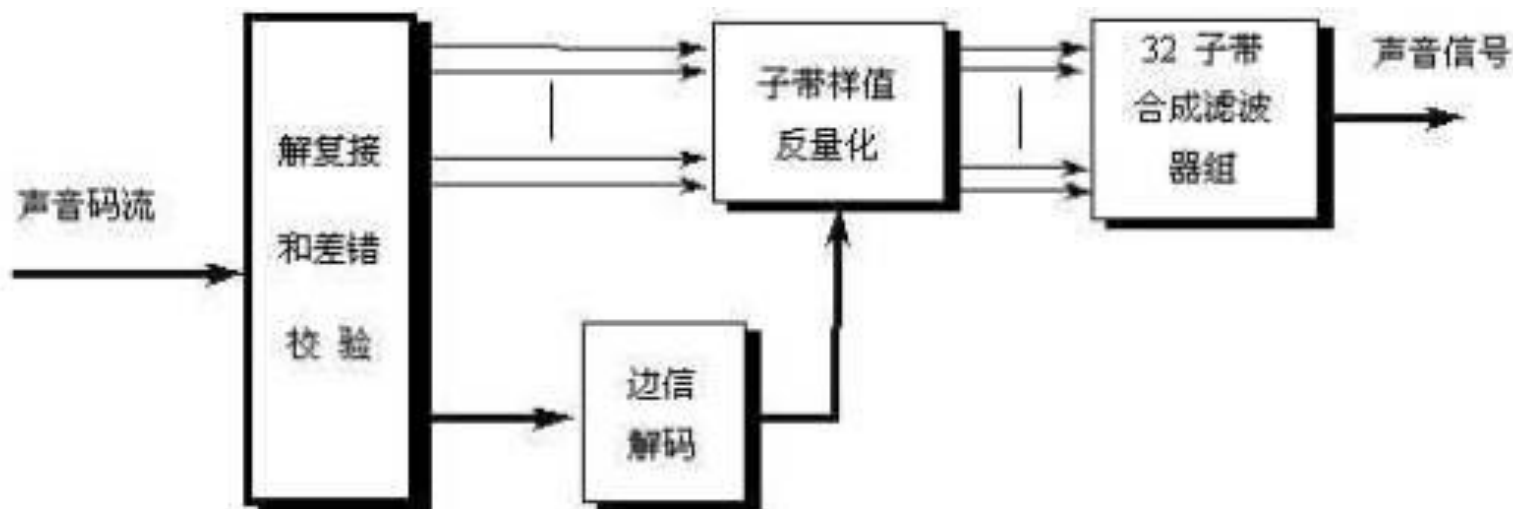
MP1/2解码

- 首先从MPEG音频第1、2层码流中将帧头信息、CRC校验字、边信息（比特分配信息和比例因子）及每子带的12个连续样值分离出来
- 重建过程用每帧各子带的比特分配信息和比例因子将样值数据扩展，综合滤波器再恢复出完整宽带的声音信号



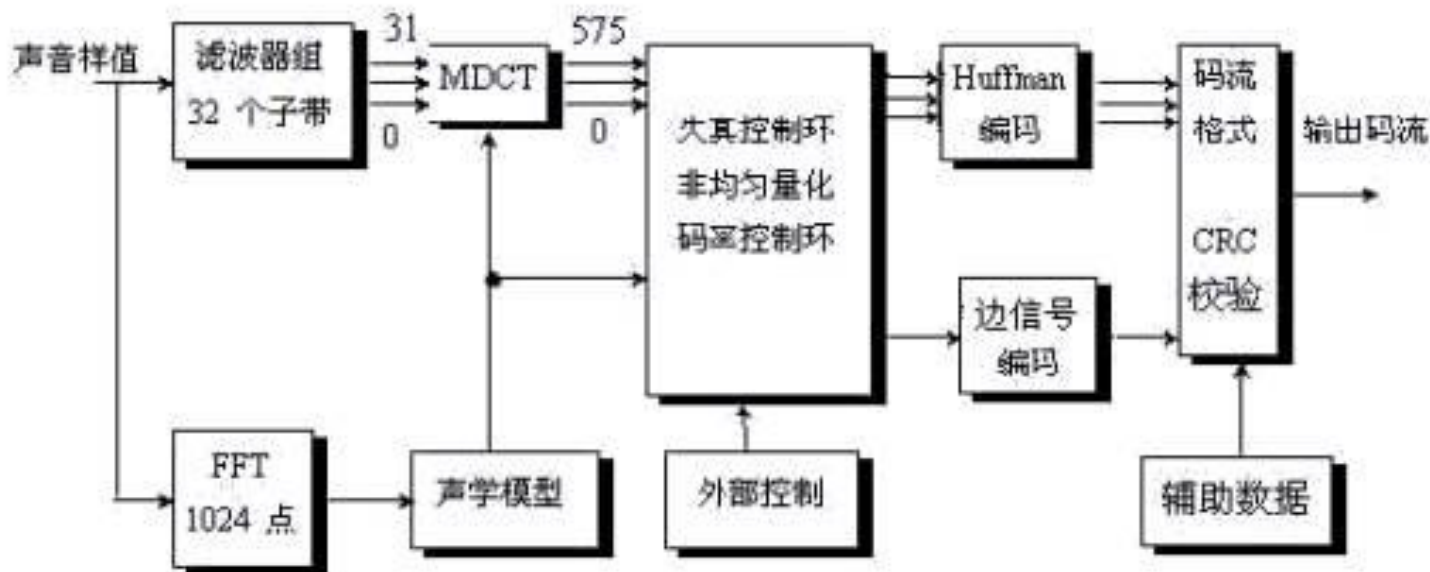
MP1/2解码

- 解码过程要求的计算能力比编码过程少得多。在第1层码的这个比例关系大约是1:2，第2层码是1:3。由于计算能力要求低和算法的直向结构，两层都可以很容易用一块专用ASIC实现。



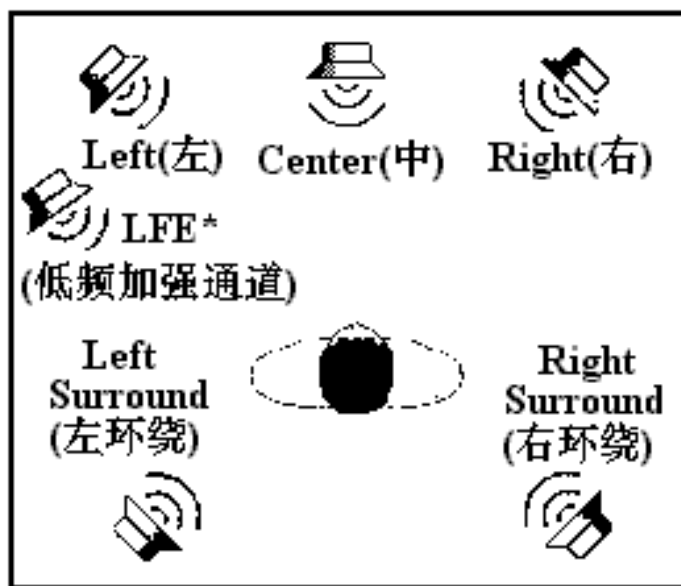
MPEG-I Audio layer 3 (mp3)

- 第3层使用的滤波器组是多相/MDCT混合滤波器组，并且使用了心理声学模型2来评估掩蔽门限，除了考虑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
- 为了增加编码增益，采用了非均匀量化和Huffman编码，并且使用了称为比特池的缓存技术来维持编码效率和使量化噪声保持在掩蔽门限以下

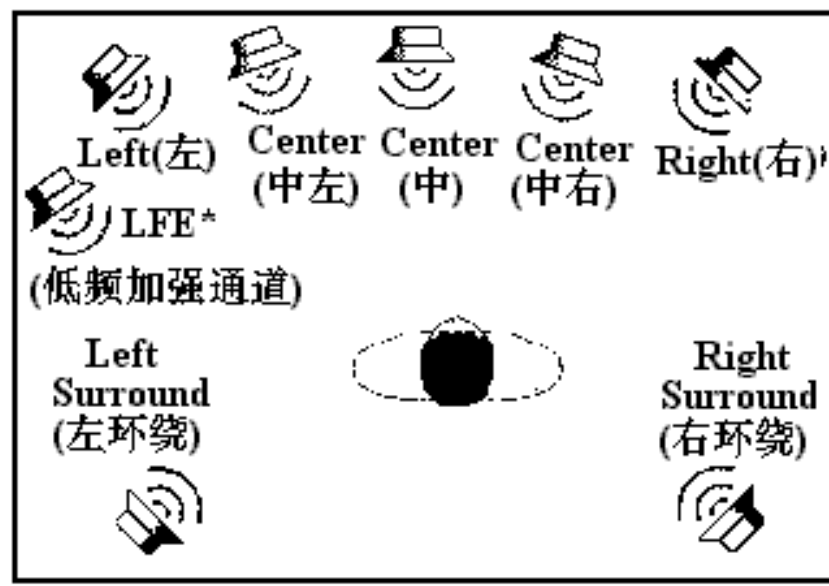


MPEG-2音频

- 增加了声道数，支持5.1声道和7.1声道的环绕声。5.1声道也称为“3/2-立体声加LFE”，其中的“.1”就是指LFE（LowFrequencyEffects，低频音效）声道。它能提供5个全带宽声道（左、右、中共3个喇叭声道，以及两个环绕声道），外加一个低频效果增强声道



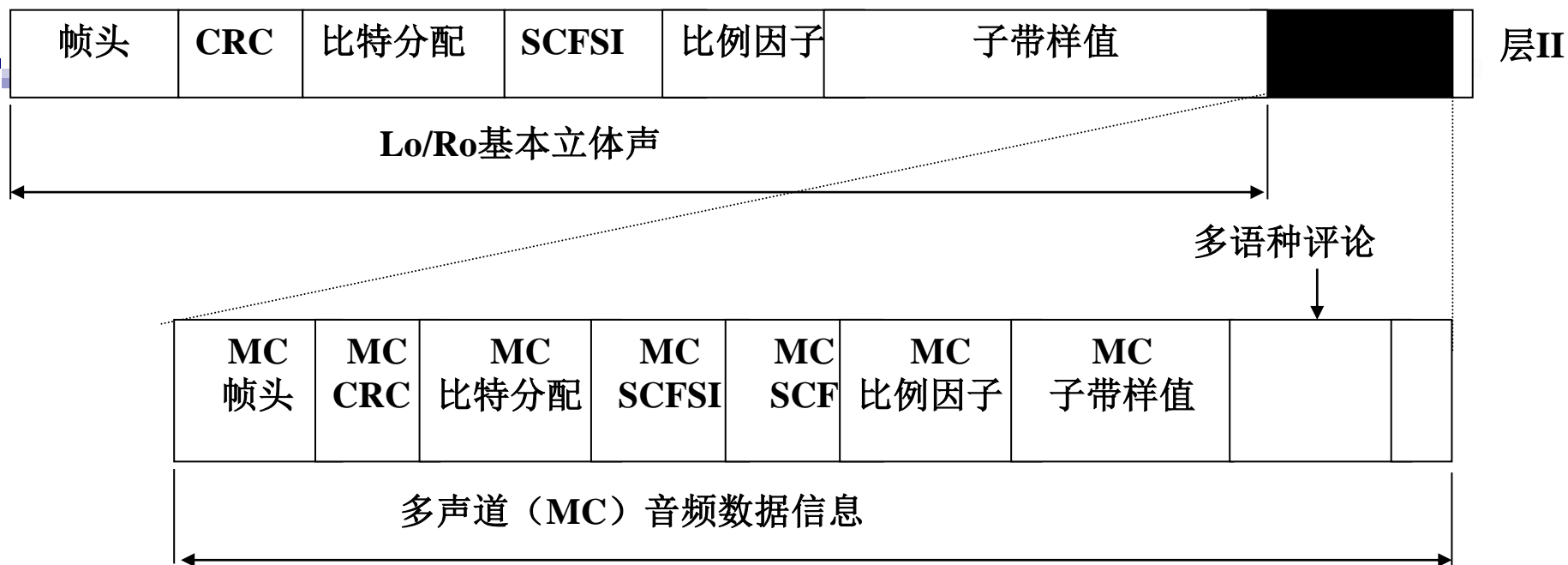
* LFE: Low Frequency Enhancement (3 Hz to 120 Hz)



* LFE: Low Frequency Enhancement (3 Hz to 120 Hz)

MPEG-2音频

- 扩展了编码器的输出速率范围，由32~384kbps扩展到8~640kbps。
- 增加了低取样和低码率
 - MPEG-1 音频:32/44.1/48kHz
 - MPEG-2增加了三种取样率:16kHz、22.05kHz、24kHz
 - 提高码率低于64kbits/s时的每个声道的声音质量。
- 和现有的MPEG-1 音频标准保持后向兼容
 - 对原有的MPEG-1两声道增加独立的环绕声道
 - 尽量保持和MPEG-1 音频语法的兼容性，MPEG-2中的主声道（左、右）仍然保持后向兼容，而环绕声道采用新的编码方法和语法



- ◆多声道中的中心声道C、左右环绕Ls、Rs及低音效果增强声道LFE等多声道扩展（MC-Extension）信息看做是MPEG-1左右声道的辅助数据而传送
- ◆多声道扩展MC部分包含了与MPEG-1声音帧结构相似的信息类型

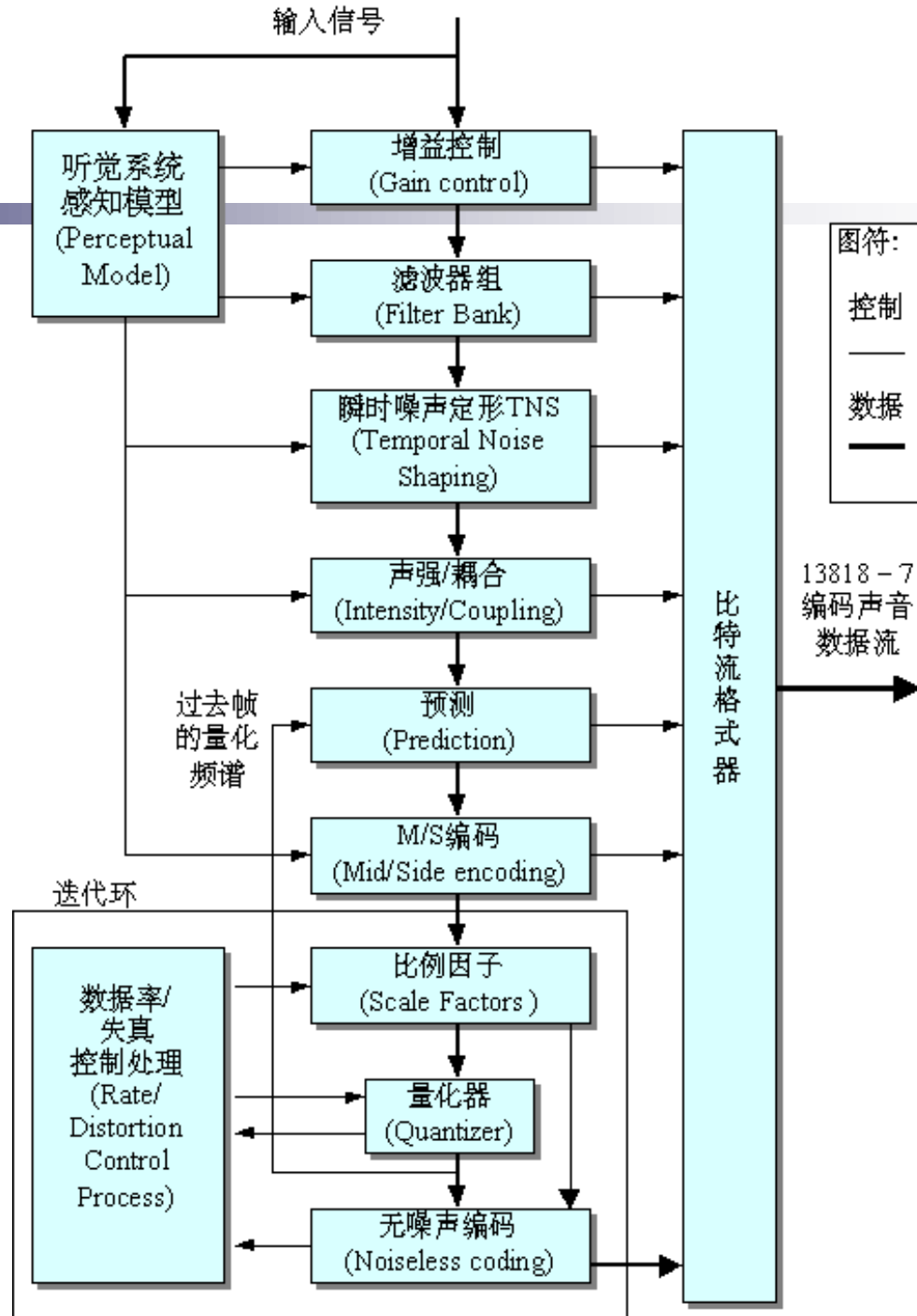
MPEG-2 AAC

- 压缩比为11:1(即每个声道的数据率为 $(44.1 \times 16) / 11 = 64 \text{ kbps}$), 5声道的总数据率为 320kbps的情况下, 很难区分还原后的声音与原始声音之间的差别
- 与MPEG-1的层2相比, MPEG-2 AAC的压缩率可提高1倍, 而且质量更高
- 与MPEG-1的层3相比, 在质量相同的条件下数据率是它的70%

MPEG-2 AAC

- 开发MPEG-2 AAC标准采用的方法与开发MPEG-1 Audio标准采用的方法不同
- 后者采用的方法是对整个系统进行标准化，而前者采用的方法是模块化的方法，把整个AAC系统分解成一系列模块，用标准化的AAC工具(advanced audio coding tools)对模块进行定义

MPEG-2 AAC



MPEG-1和-2的声音数据规格

参数名称	Linear PCM	Dolby AC-3	MPEG-2 Audio	MPEG-1 Audio
采用频率	48/96 kHz	32/44.1/48 kHz	16/22.05/24/32/44.1/48 kHz	32/44.1/48 kHz
样本精度 (每个样本的比特数)	16/20/24	压缩 (16 bits)	压缩 (16 bits)	16
最大数据传输率	6.144 Mbit/s	448 kbps	8~640 kbps	32~448 kbps
最大声道数	8	5.1	5.1/7.1	2