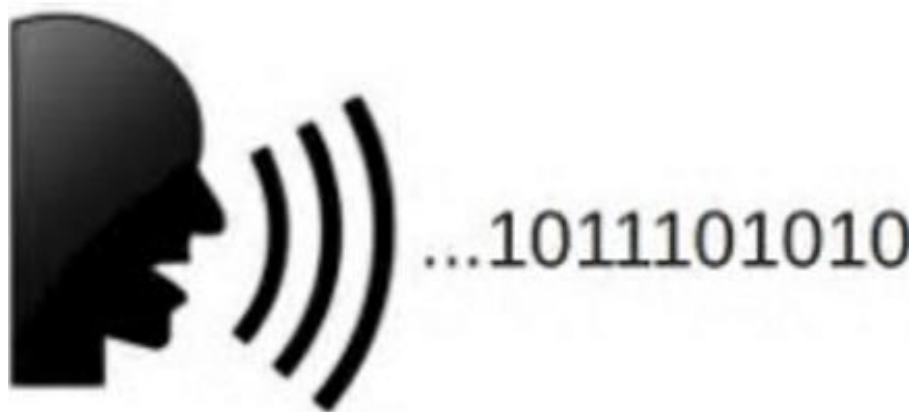


## 第3章 数字声音编码



## 3.1 简介

3.1.1 声音简介

3.1.2 声音信号数字化

3.1.3 声音质量的MOS评分标准

## 3.2 声音的编码

3.2.1 脉冲编码调制 (PCM) 及其在通信中的应用

3.2.2 增量调制 (DM)与自适应增量调制(ADM)

3.2.3 自适应差分脉冲编码调制(ADPCM)

- 自适应脉冲编码调制 (APCM)

- 差分脉冲编码调制 (DPCM)

- 自适应差分脉冲编码调制(ADPCM)

- G.722 SB-ADPCM编译码器

3.2.4 线性预测编码(LPC)及GSM声音简介



# 目录

---

## 3.1 简介

3.1.1 声音简介

3.1.2 声音信号数字化

3.1.3 声音质量的MOS评分标准

## 3.2 声音的编码

3.2.1 脉冲编码调制 (PCM) 及其在通信中的应用

3.2.2 增量调制 (DM)与自适应增量调制(ADM)

3.2.3 自适应差分脉冲编码调制(ADPCM)

- 自适应脉冲编码调制 (APCM)

- 差分脉冲编码调制 (DPCM)

- 自适应差分脉冲编码调制(ADPCM)

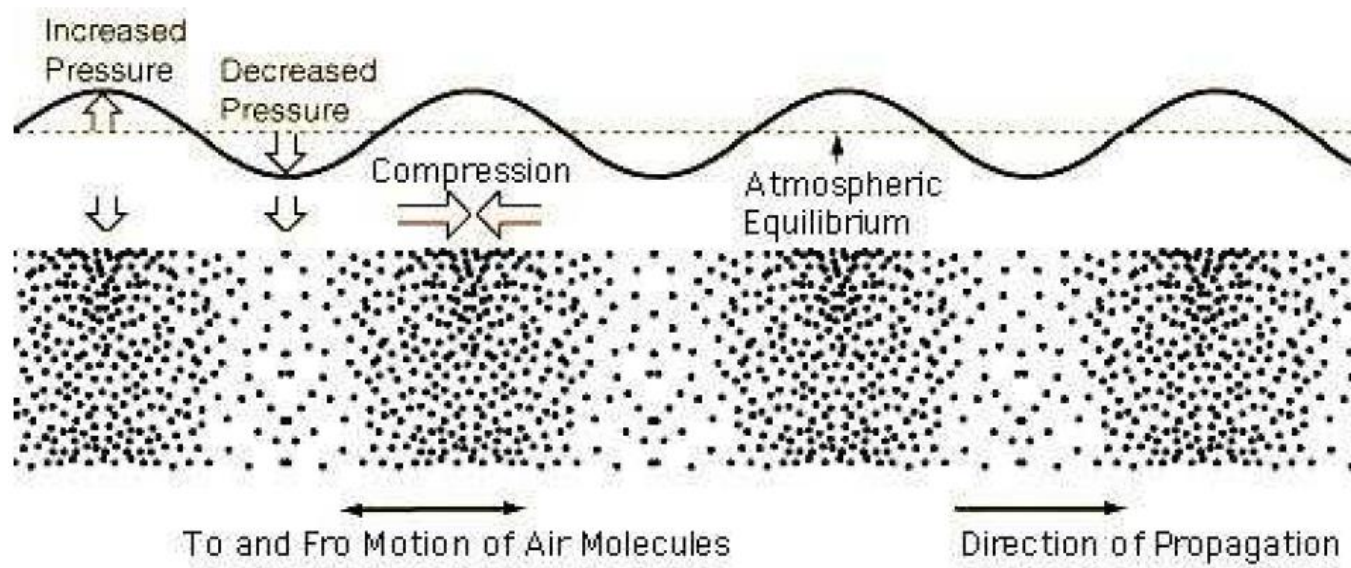
- G.722 SB-ADPCM编译码器

3.2.4 线性预测编码(LPC)及GSM声音简介

## 3.1.1 声音简介

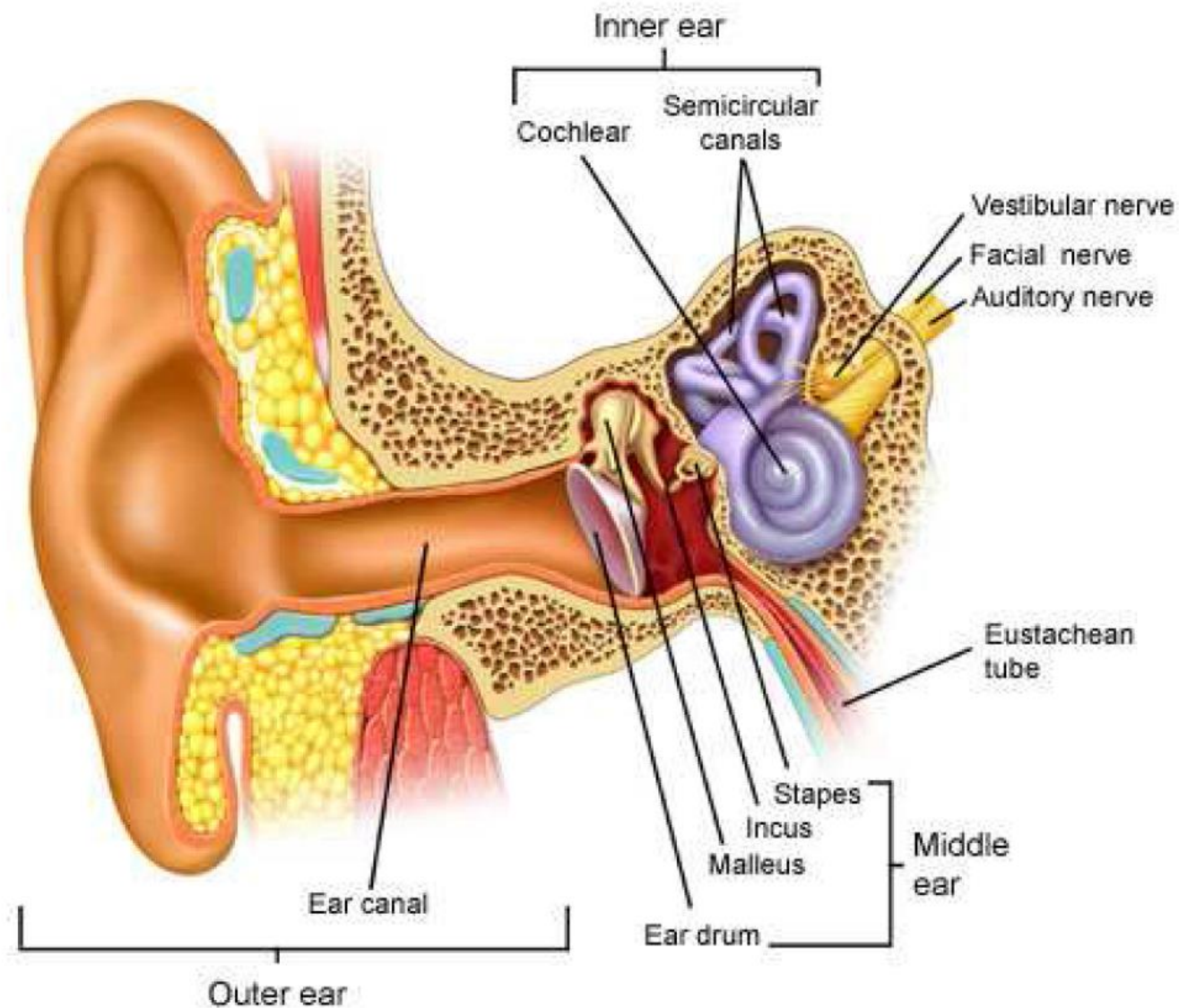
### ■ 什么是声音？

- 声音是听觉器官对声波的感知，而声波是通过空气或其他媒体传播的连续振动



- 在20度的空气中，声波的传播速度是343.2m/s

### 3. 1. 1 声音简介



### 3.1.1 声音简介

- 声波具有普通波所具有的特性，例如反射 (reflection)、折射 (refraction) 和衍射 (diffraction) 等

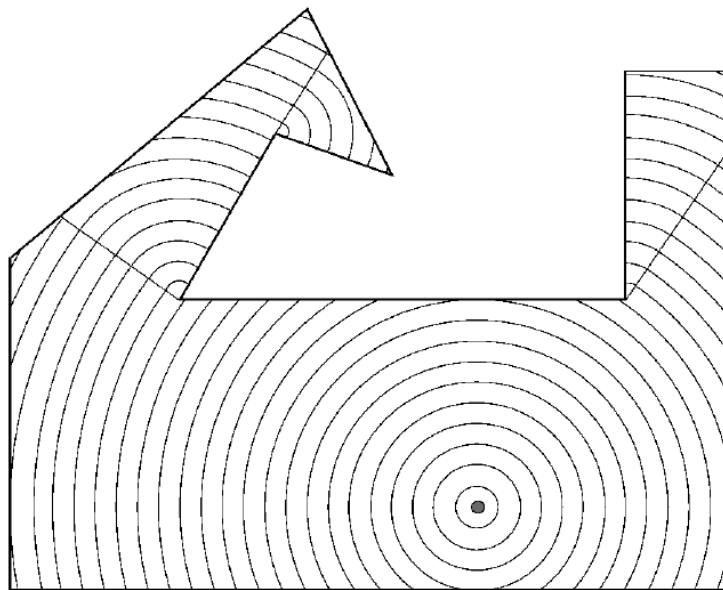


Figure 11.2: Waves can even bend around corners, due to *diffraction*. A top-down view of a room is shown. At each of the three interior corners, the propagating wavefront expands around it.

### 3.1.1 声音简介

- 声音的强弱体现在声波压力的大小上，声波压力可用decibels(dB)来衡量。

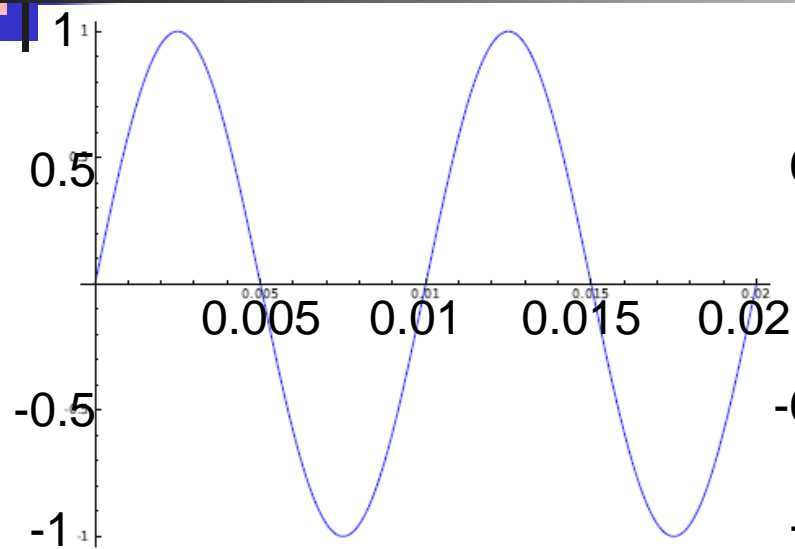
$$N_{db} = 20 * \log_{10}(p_e/p_r)$$

压力的峰值

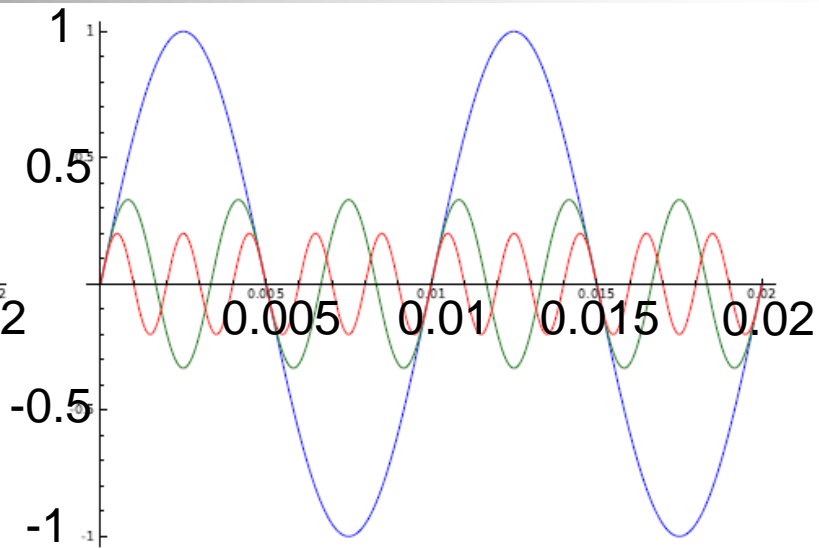
$2 \times 10^{-5} \text{pa}$

- 音调的高低体现在声音的频率上。一个纯音可以用正弦波来表示，只有一个对应的频率；更复杂的声音可以由一系列纯音组成。傅立叶分析可以把一个周期函数分解成多个正弦波，对分析声源特性非常重要。

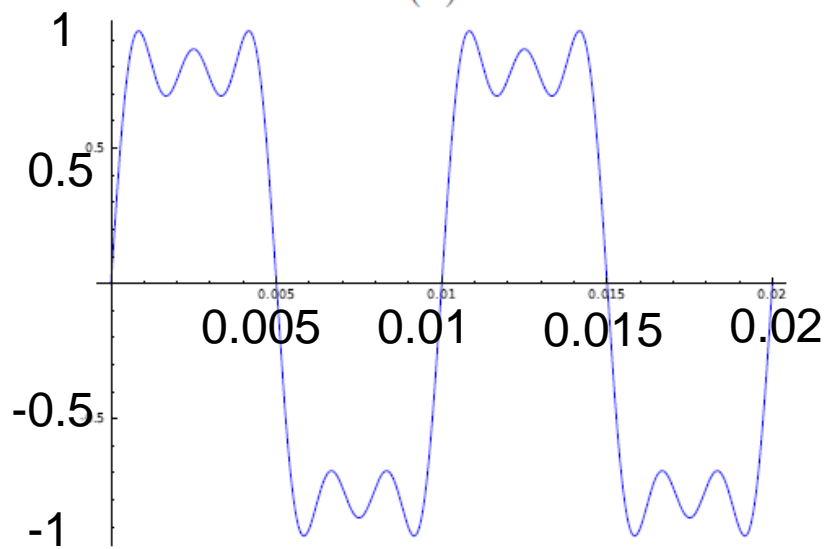
# 声波示例



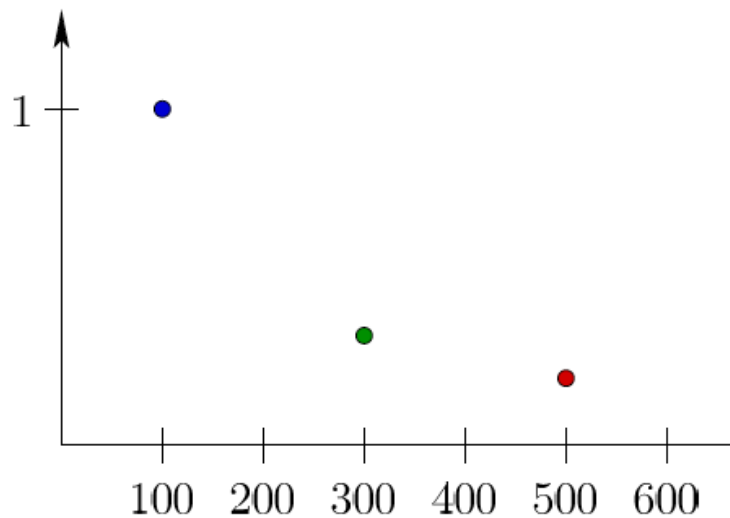
(a)



(b)



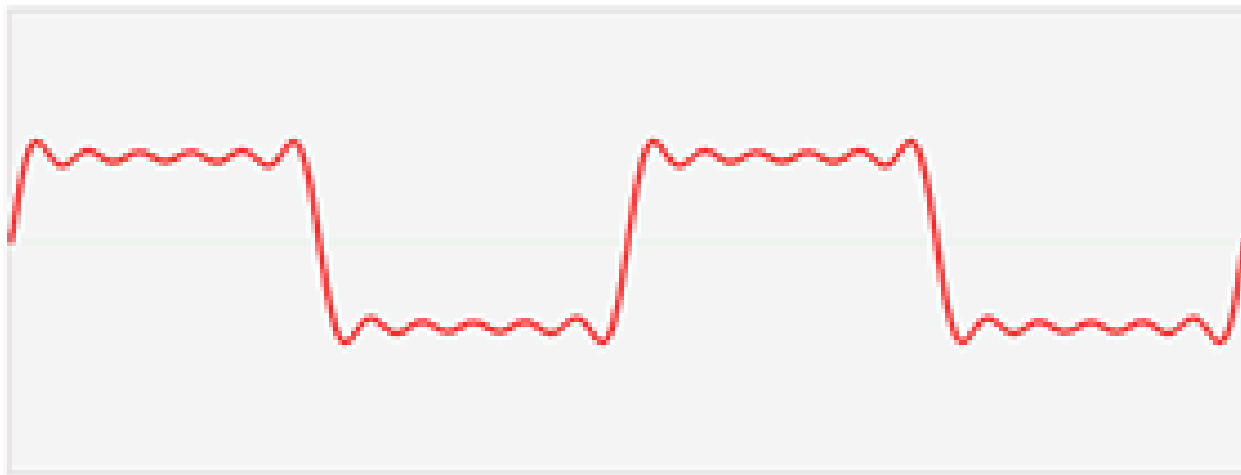
(c)



(d)



# 声波示例



### 3.1.1 声音简介

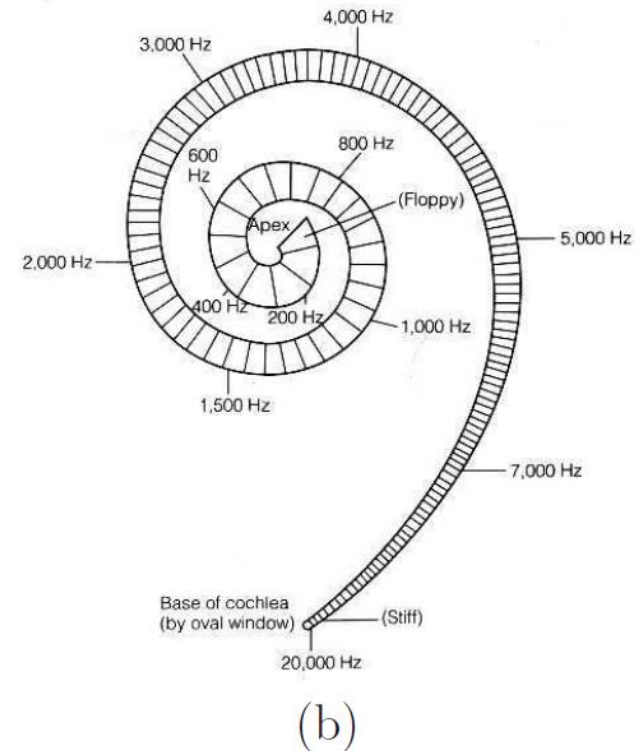
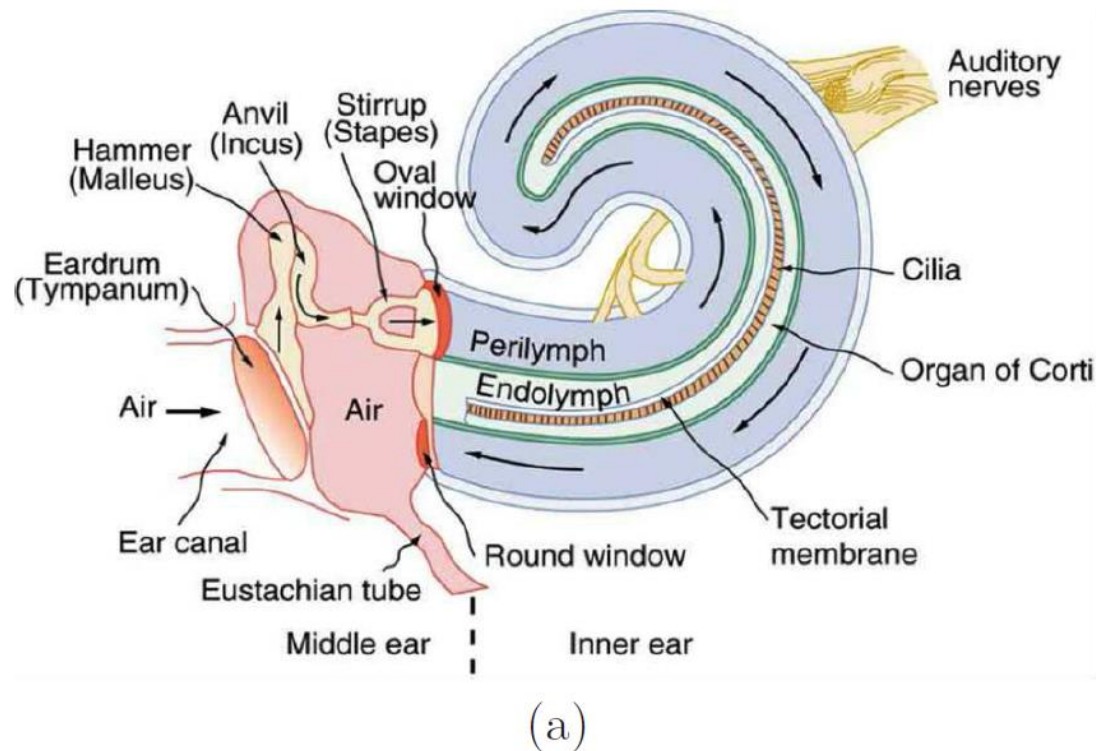


Figure 11.5: The operation of the cochlea: (a) The perilymph transmits waves that are forced by the oval window through a tube that extends the length of the cochlea and back again, to the round window. (b) Because of varying thickness and stiffness, the central spine (basilar membrane) is sensitive to particular frequencies of vibration; this causes the mechanoreceptors, and ultimately auditory perception, to be frequency sensitive.



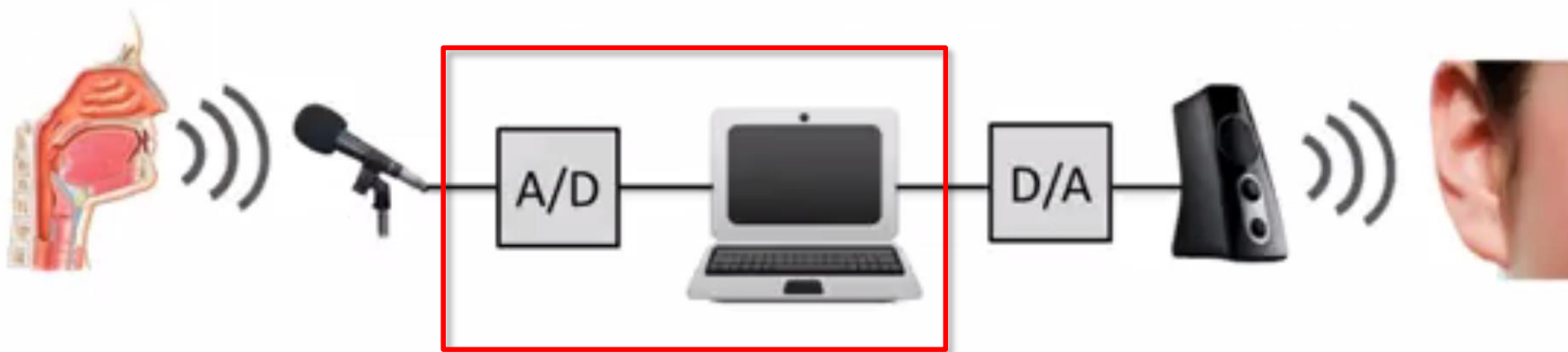
## 3.1.1 声音简介

---

### ■ 声音的频率

- 声音(audio): 20~ 20 000Hz
- 高保真声音(high-fidelity audio): 10 ~ 20 000 Hz
- 话音(speech): 300~3000/3400 Hz
- 亚音/次音(subsonic):  $< 20$  Hz
- 超声(ultrasonic):  $> 20\ 000$  Hz

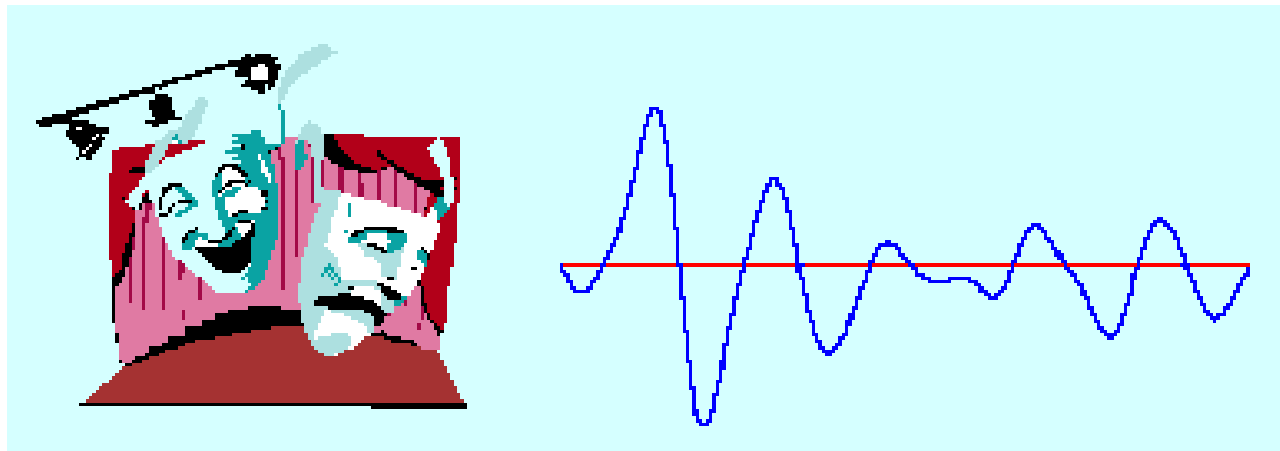
### 3.1.1 声音简介



声音的产生、传播、处理、感知示例

## 3.1.1 声音简介

- 声音用电压信号表示时，声音信号在时间和幅度上都是连续的模拟信号

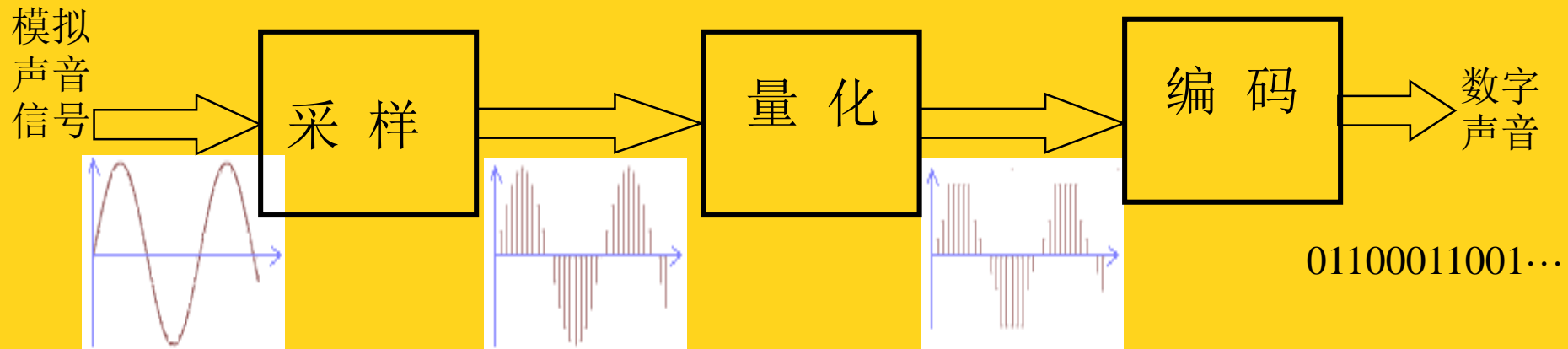


## 3.1.2 声音信号数字化

### ■ 数字化的概念

#### ➤ 采样和量化

- 连续时间的离散化通过采样来实现。
  - 均匀采样(uniform sampling): 每隔相等的一段时间采样一次
- 连续幅度的离散化通过量化(quantization)来实现。
  - 线性量化: 把信号的强度划分成等间隔的多个小段





## 3.1.2 声音信号数字化

### ■ 采样频率

- 采样频率由根据奈奎斯特-香农采样定理(Nyquist-Shannon Sampling Theorem)确定
  - 无损数字化(lossless digitization): 采样频率不低于声音信号最高频率的两倍, 就能把以数字表达的声音还原成原来的声音。
- 假设被采样信号的最高频率为 $f_{\max}$ , 则采样定律表示为

$$f_s \geq 2f_{\max}$$



## 3.1.2 声音信号数字化

### ■ 采样精度

- 度量声音波形幅度的精确程度，用每个声音样本的位数(即bps)表示
  - 精度是在模拟信号数字化过程中度量模拟信号的最小单位，也称量化阶(quantization step size)。
- 样本位数的大小影响到声音的质量
  - 位数越多，声音质量越高，所需存储空间也越多。



## 3.1.2 声音信号数字化

- 采样精度也可用信号噪声比表示，简称为信噪比(signal-to-noise ratio, SNR),

$$SNR = 10 \log_{10} \left[ \frac{(V_{signal})^2}{(V_{noise})^2} \right] = 20 \log_{10} \left( \frac{V_{signal}}{V_{noise}} \right) \quad \text{单位为分贝 (dB)}$$

$V_{signal}$ : 信号电压

$V_{noise}$ : 量化噪声电压(模拟信号的采样值和与它最接近的数字数值之间的差值)

## 3.1.2 声音信号数字化

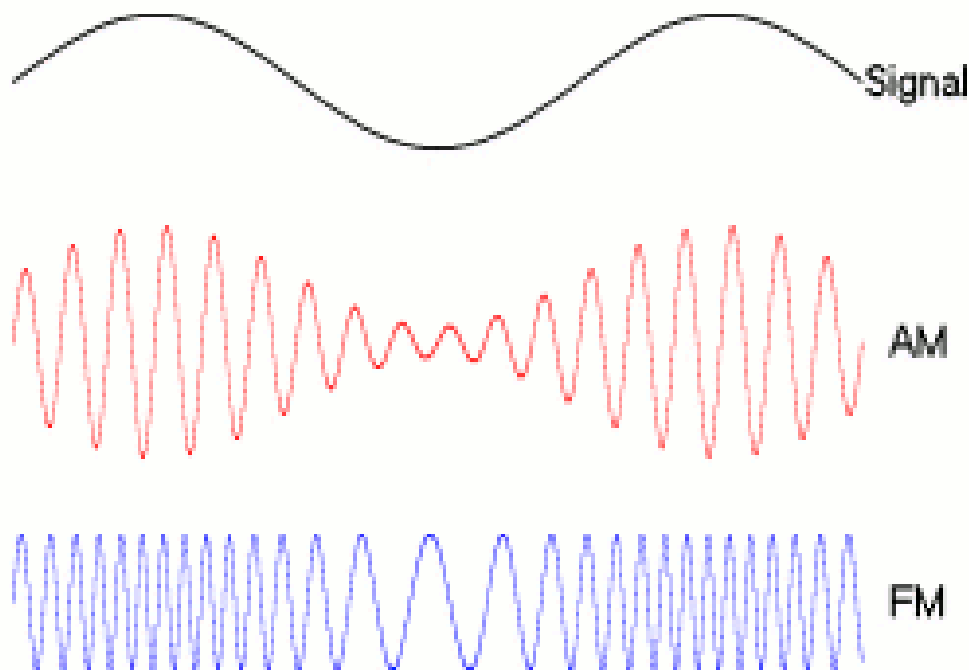
### ■ 声音质量和数据率——质量度量

质量	采样频率 (kHz)	样本精度 (bit/s)	单道声/立 体声	(未压缩的) 数据率(kb/s)	频率范围 (Hz)
电话*	8	8	单道声	64.0	200~3 400
AM	11.025	8	单道声	88.2	20~ 15000
FM	22.050	16	立体声	705.6	50~7000
CD	44.1	16	立体声	1411.2	20~ 20000
DAT	48	16	立体声	1536.0	20~ 20000

\*AM: amplitude modulation  
FM: frequency modulation

CD: compact disc audio  
DAT: digital audio tape

## 3.1.2 声音信号数字化



中央人民广播电台节目频率分布表

2009年1月1日00:00起执行

语言	播音时间	频率{千赫}	服务区	备注
中一	00:00-24:00	106.1	北京	调频
中一	03:55-01:35	639	华北	中波
中一	03:55-01:35	6030	华北	



### 3.1.3 声音质量的衡量方法

#### ■ 声音质量的衡量方法

- 声音带宽法
  - 等级由高到低依次是DAT、CD、FM、AM和数字电话
- 客观质量度量
  - 信噪比（SNR）
- 主观质量度量
  - 主观平均分（MOS）

# 3.1.3 声音质量的MOS评分标准

## ■ 主观平均分(mean opinion score, MOS)

- 对声音主观质量度量比较通用的标准是5分制

声音质量MOS评分标准

分数	质量等级	失真级别
5	优(Excellent)	无察觉
4	良(Good)	(刚)察觉但不讨厌
3	中(Fair)	(察觉)有点讨厌
2	差(Poor)	讨厌但不反感
1	劣(Bad)	极讨厌(令人反感)



# 目录

---

## 3.1 简介

3.1.1 声音简介

3.1.2 声音信号数字化

3.1.3 声音质量的MOS评分标准

## 3.2 声音的编码

3.2.1 脉冲编码调制 (PCM) 及其在通信中的应用

3.2.2 增量调制 (DM)与自适应增量调制(ADM)

3.2.3 自适应差分脉冲编码调制(ADPCM)

- 自适应脉冲编码调制 (APCM)

- 差分脉冲编码调制 (DPCM)

- 自适应差分脉冲编码调制(ADPCM)

- G.722 SB-ADPCM编译码器

3.2.4 线性预测编码(LPC)及GSM声音简介

## 3.2.1 脉冲编码调制(PCM)

### ■ 脉冲编码调制(Pulse Code Modulation, PCM)

概念上最简单、理论上最完善、最早研制成功、使用最为广泛、数据量最大的编码系统

- 输入是模拟信号，输出是PCM样本。
- 防失真滤波器：低通滤波器，用来滤除声音频带以外的信号
- 波形编码器：可理解为采样器
- 量化器：可理解为“量化阶大小(step-size)”生成器或者称为“量化间隔”生成器



## 3.2.1 脉冲编码调制

### ■ PCM实际上是模拟信号数字化

模拟声音数字化的两个步骤：

- 第一步是采样，就是每隔一段时间间隔读一次声音的幅度
- 第二步是量化，就是把采样得到的声音信号幅度转换成数字值

### ■ 量化的方法

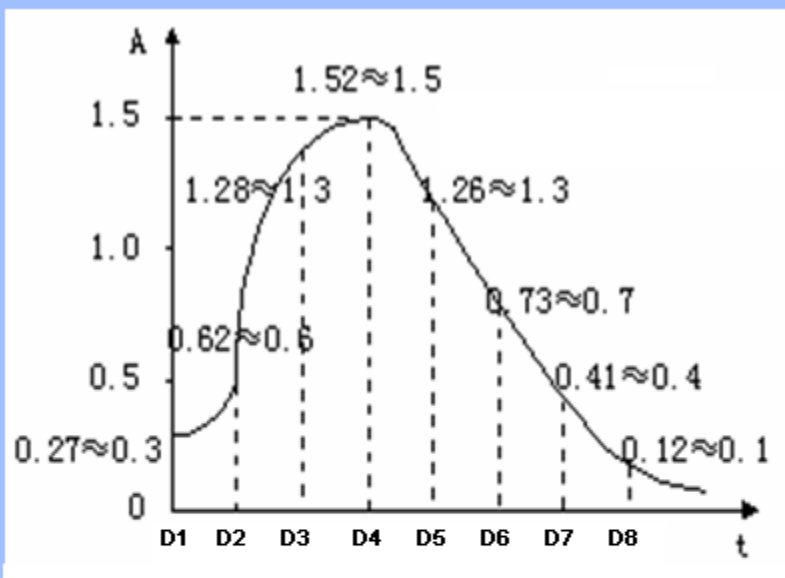
- 主要有均匀量化和非均匀量化



## 3.2.1 脉冲编码调制

### ■ 均匀量化

- 采用相等的量化间隔对采样得到的信号作量化。
- 量化误差(量化噪声): 量化后的样本值 $Y$ 和原始值 $X$ 的



[0,1.5]分成15个区间, 间隔长度0.1, 形成16个量化级。

样本	量化级	二进制编码	编码信号
D <sub>1</sub>	3	0011	
D <sub>2</sub>	6	0110	
D <sub>3</sub>	13	1101	
D <sub>4</sub>	15	1111	
D <sub>5</sub>	13	1101	
D <sub>6</sub>	7	0111	
D <sub>7</sub>	4	0100	
D <sub>8</sub>	1	0001	

## 3.2.1 脉冲编码调制

### ■ 非均匀量化

- 大的输入信号采用大的量化间隔，小的输入信号采用小的量化间隔
- 可在满足精度要求的情况下用较少的位数来表示
- 声音数据还原时，采用相同的规则

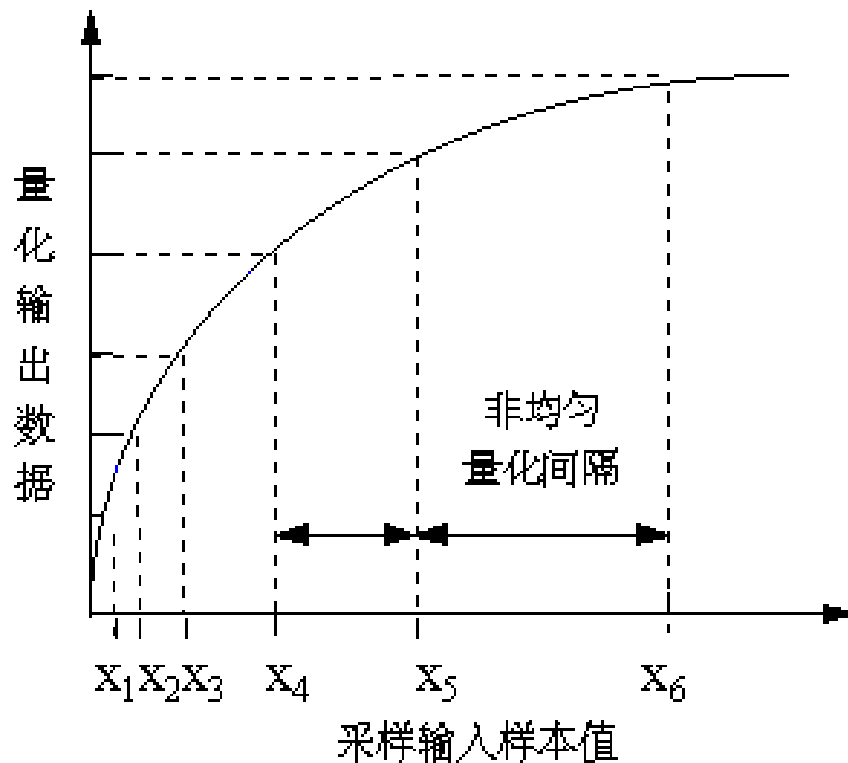


图3-5 非均匀量化

## 3.2.1 脉冲编码调制

- 采样输入信号幅度和量化输出数据之间定义了两种对应关系
  - $\mu$ 律压扩算法
  - A律压扩算法

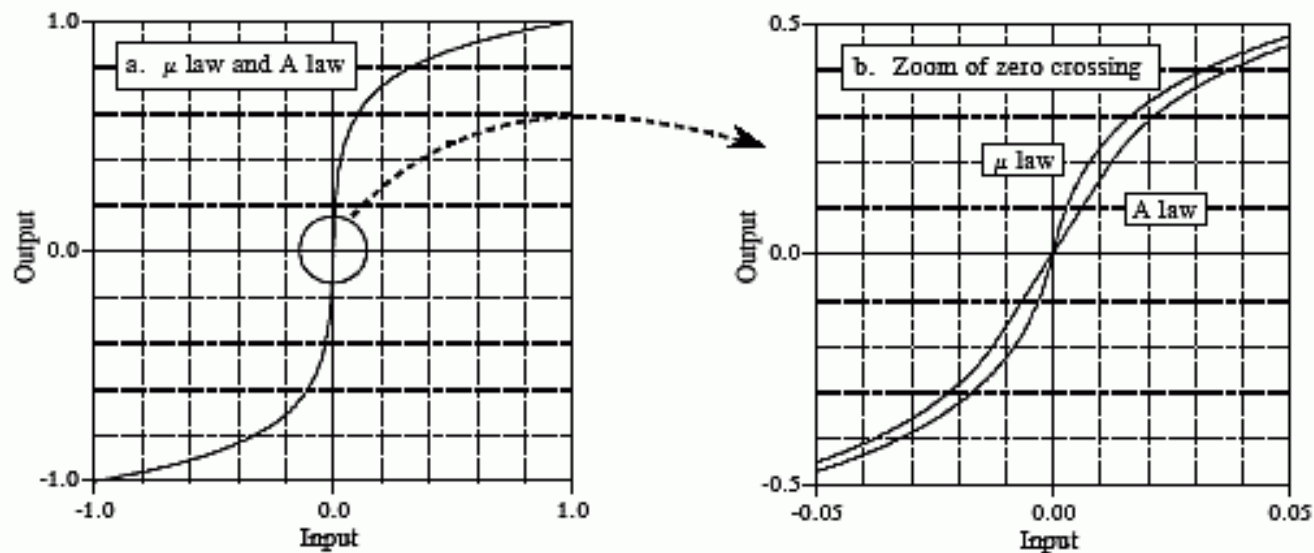


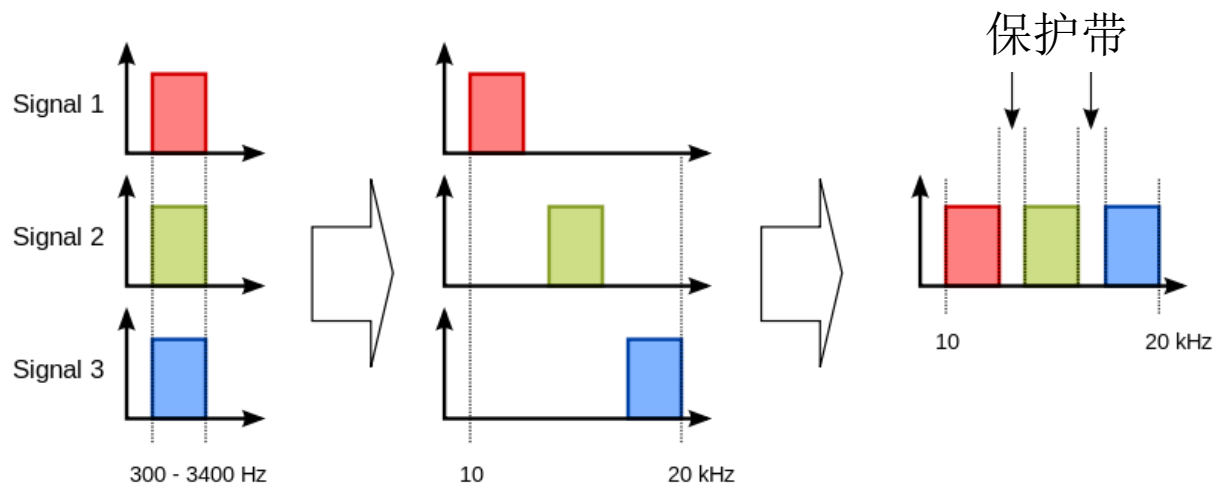
FIGURE 22-7

Companding curves. The  $\mu$ 255 law and "A" law companding curves are nearly identical, differing only near the origin. Companding increases the amplitude when the signal is small, and decreases it when it is large.

## 3.2.1 PCM在通信中的应用

### ■ 频分多路复用

- 在一条通信线路上使用不同频段同时传送多个独立信号的通信方法
- 模拟载波通信的主要方法



每一个信号的频率平移到一个特定的频率范围内

## 3.2.1 PCM在通信中的应用

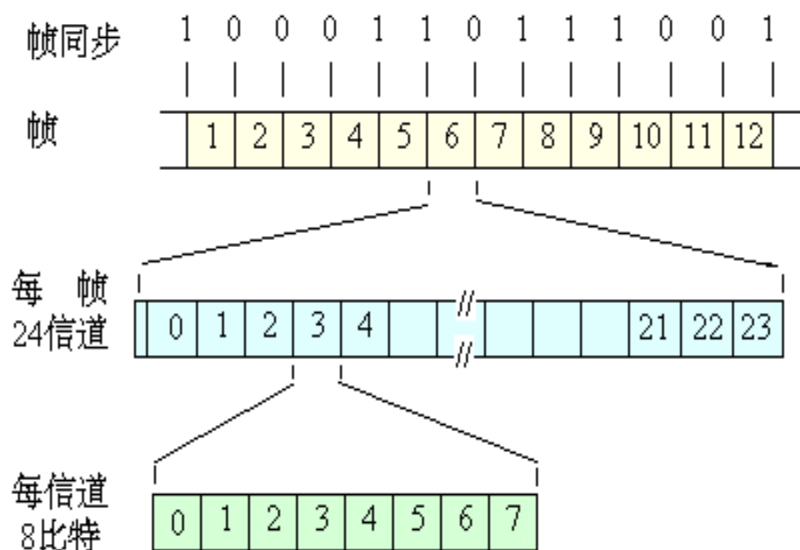
### ■ 时分多路复用

- 在一条通信线路上使用不同时段“同时”传送多个独立信号的通信方法
- 核心思想：将时间分成等间隔的时段，为每对用户指定一个时间间隔，每个间隔传输信号的一部分
- 例如，话音信号的采样频率 $f=8000\text{ Hz/s}$ ，采样周期 $=125\text{ }\mu\text{s}$ ，这个时间称为1帧(frame)。在这个时间里可容纳的话路数有两种规格
  - 24路制
  - 30路制

## 3.2.1 PCM在通信中的应用

### ■ 24路制的重要参数如下：

- 每秒钟传送8000帧，每帧  $125\mu\text{s}$
- 12帧组成1复帧(用于同步)。
- 每帧由24个时间片(信道)和1位同步位组成
- 每个信道每次传送8位代码，1帧有  $24 \times 8 + 1 = 193$  位(位)
- 数据传输率  $R =$
- 每一个话路的数据传输率 =



24路PCM的帧结构

## 3.2.1 PCM在通信中的应用

### ■ 30路制的重要参数如下:

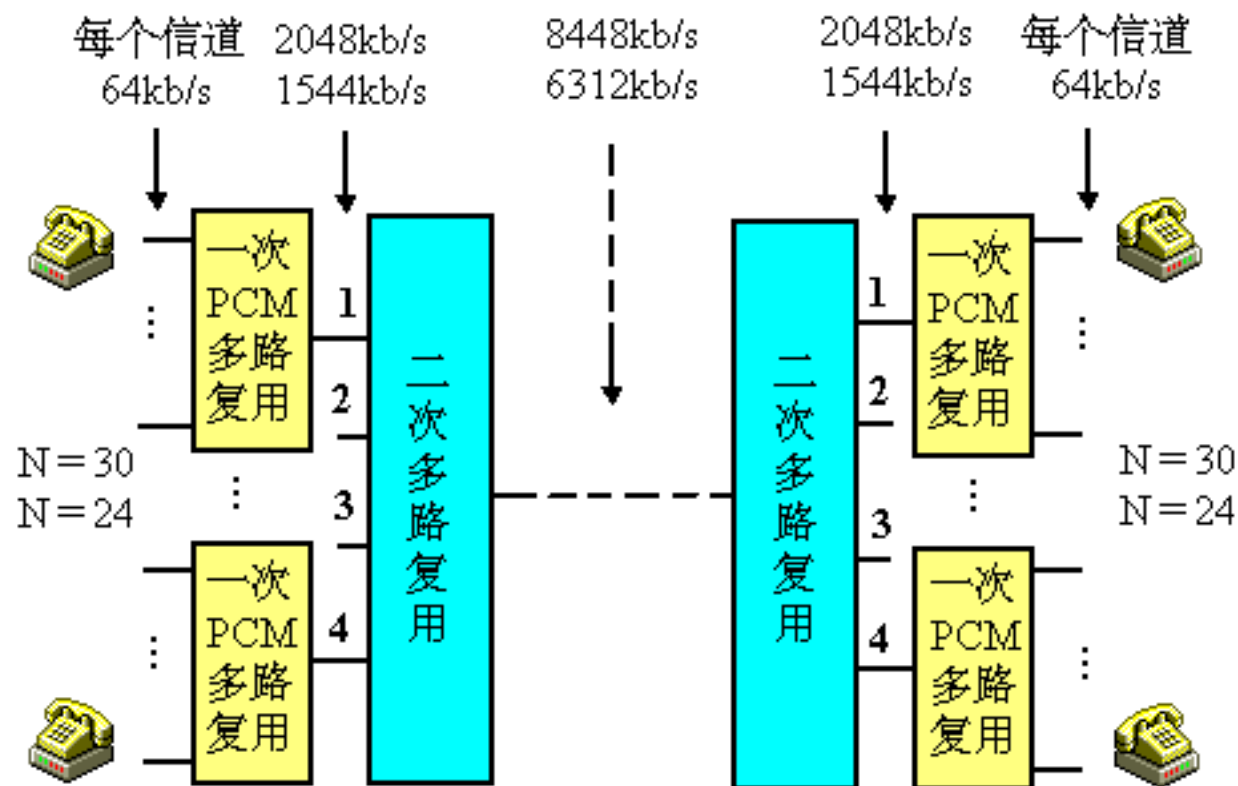
- 每秒钟传送8000帧, 每帧125  $\mu\text{s}$
- 16帧组成1复帧(用于同步)
- 每帧由32个时间片(信道)组成
- 每个信道每次传送8位代码
- 数据传输率:  $R =$
- 每一个话路的数据传输率 =

### ■ 线路利用率

- 当信道无数据传输时仍给那个信道分配时间槽, 因此线路利用率较低
- 统计时分多路复用技术(statistical time division multiplexing, STDM) 按照每个传输信道的传输需要来分配时间间隔的时分多路复用技术, 可提高传输线路的效率

## 3.2.1 PCM在通信中的应用

### 数字通信线路的数据传输率



二次复用示意图



## 3.2.1 PCM在通信中的应用

- T1/E1, T2/E2, T3/E3, T4/E4和T5/E5的数据传输率

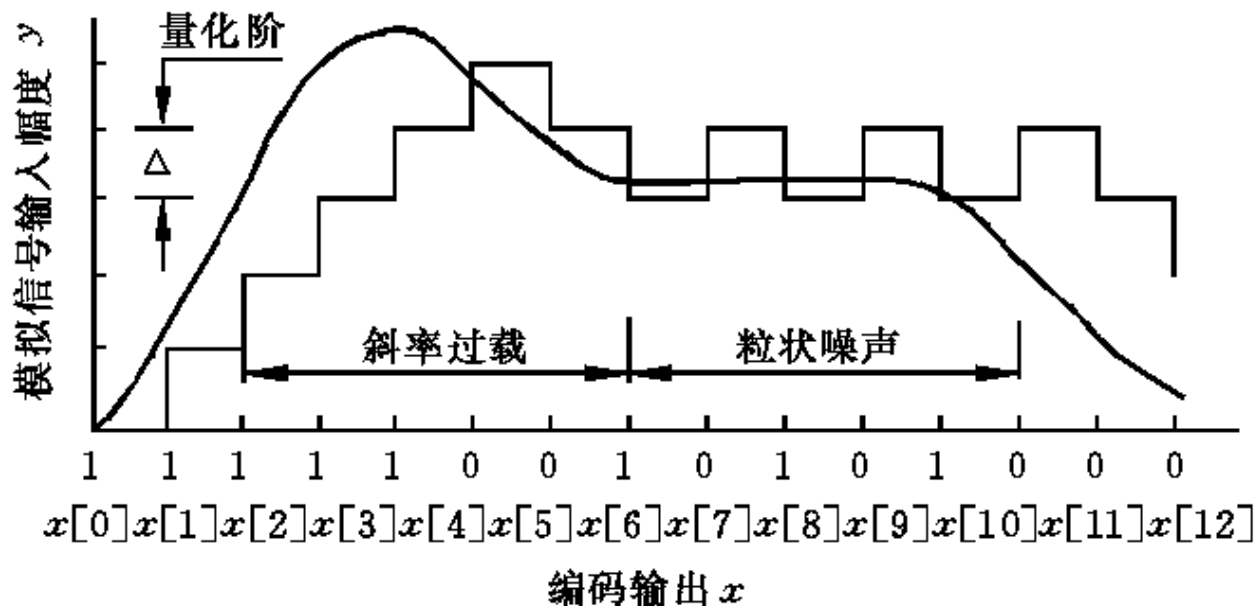
表 3-3 多次复用的数据传输率(注)

国别	数字网络等级	T1/E1	T2/E2	T3/E3	T4/E4	T5/E5
美国	64 kb/s话路数	24	96	672	4032	-
	总传输率(Mb/s)	1.544	6.312	44.736	274.176	-
	数字网络等级	1	2	3	4	5
欧洲	64 kb/s话路数	30	120	480	1920	7680
	总传输率(Mb/s)	2.048	8.448	34.368	139.264	560.000
日本	64 kb/s话路数	24	96	480	1440	-
	总传输率(Mb/s)	1.544	6.312	32.064	97.728	-

## 3.2.2 增量调制与自适应增量调制

### ■ 增量调制 (Delta Modulation, DM)

- 也称 $\Delta$ 调制，是一种预测编码技术
- 对实际的采样信号与预测的采样信号之差的极性进行编码。如果极性为“正”，用“1”表示；反之用“0”表示，或相反
- 由于DM编码只用1位对话音信号编码，所以称为“1位系统”
  - 比较：PCM是对采样信号的整个幅度进行量化编码



## 3.2.2 增量调制与自适应增量调制

- 自适应增量调制(Adaptive Delta Modulation, ADM)
  - 根据输入信号斜率的变化自动调整量化阶 $\Delta$ 的大小，以减少斜率过载和粒状噪声。
  - 检测到斜率过载时增大量化阶 $\Delta$ ，斜率减小时降低量化阶 $\Delta$ 
    - 宋(Song)：每当输出不变时量化阶增大50%；每当输出值改变时，量化阶减小50%
    - 格林弗基斯(Greefkes)的连续可变斜率增量调制(CVSD)：如果连续可变斜率增量调制器的输出连续出现三个相同值时，量化阶加一个大的增量，反之，就加一个小的增量。



### 3.2.3 自适应脉冲编码调制

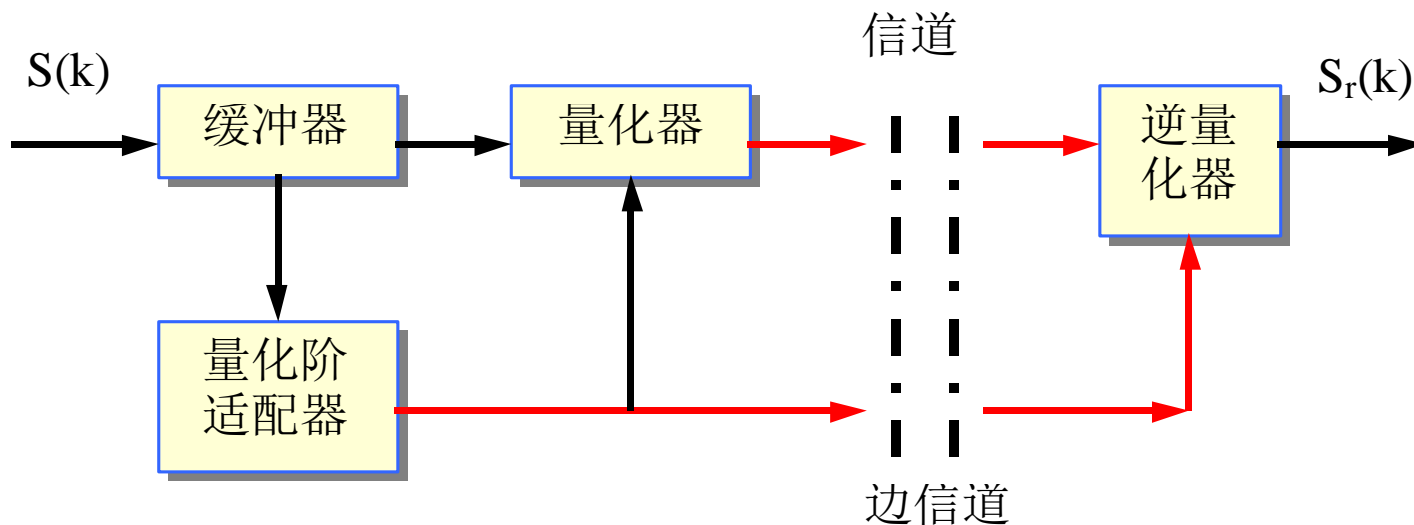
#### ■ 自适应脉冲编码调制(Adaptive Pulse Code Modulation, APCM)

- 根据输入信号幅度大小来改变量化阶大小的一种波形编码技术
- 自适应
  - 瞬时自适应：量化阶大小每隔几个样本就改变
  - 音节自适应：量化阶大小在较长时间里发生变化

### 3.2.3 自适应脉冲编码调制

➤ 前向自适应(forward adaptation)

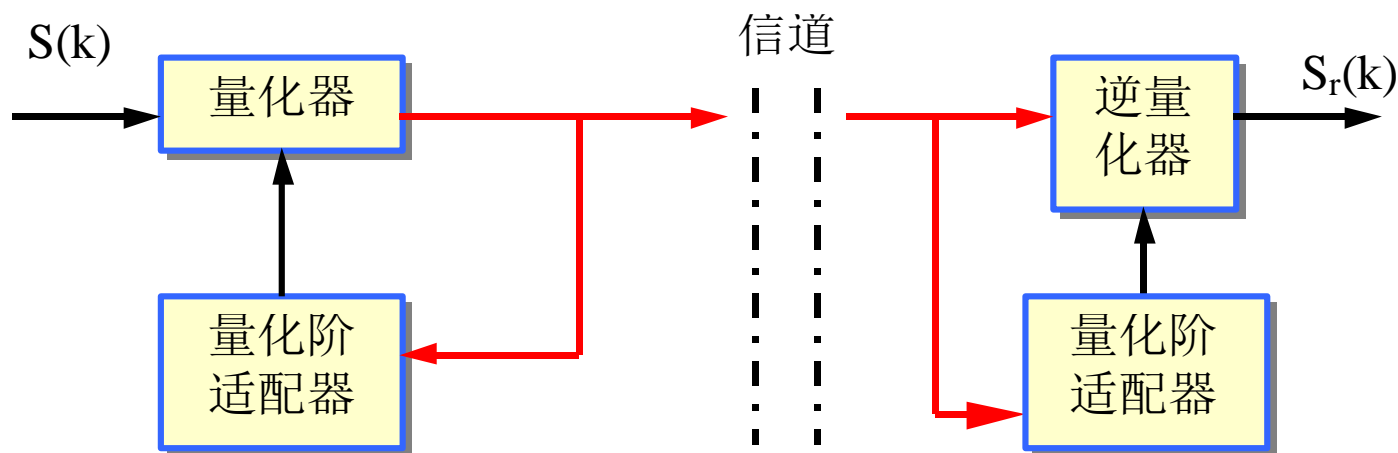
- 根据未量化的样本值的均方根值来估算输入信号的电平，以此来确定量化阶的大小，并对其电平进行编码作为边信息(side information)传送到接收端



(a)前向自适应APCM

### 3.2.3 自适应脉冲编码调制

- 后向自适应(backward adaptation)
  - 从量化器刚输出的过去样本中提取量化阶信息。由于后向自适应能在发收两端自动生成量化阶，所以它不需要传送边信息。



(b)后向自适应APCM



### 3.2.3 差分脉冲编码调制

- 差分脉冲编码调制(Differential Pulse Code Modulation, DPCM)
  - 利用样本间存在的信息冗余来进行数据压缩
  - 基本思想：根据过去的样本去估算下一个样本信号的幅度大小，这个值称为预测值，然后对实际信号值与预测值之差进行量化编码，从而就减少了表示每个样本信号的位数



### 3.2.3 差分脉冲编码调制

- 差分脉冲编码调制(Differential Pulse Code Modulation, DPCM)

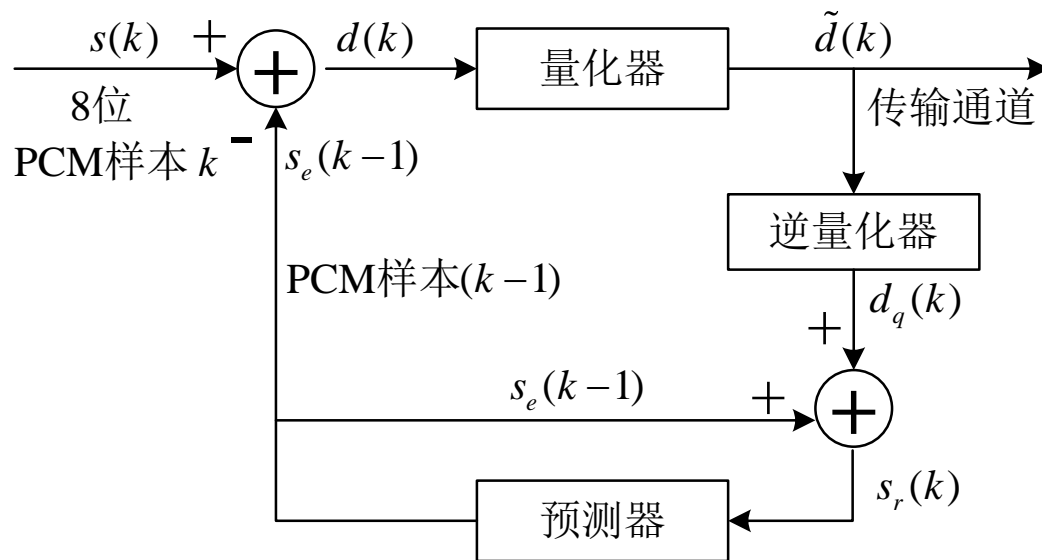
- 它与脉冲编码调制(PCM)不同的是, PCM是直接对采样信号进行量化编码, 而DPCM是对实际信号值与预测值之差进行量化编码。
- 存储或者传送的是差值而不是幅度绝对值, 这就降低了传送或存储的数据量, 可适应大范围变化的输入信号。



### 3.2.3 差分脉冲编码调制

#### ■ DPCM原理

- 差分信号 $d(k)$ : 离散输入信号 $s(k)$ 和预测器输出的估算值 $s_e(k-1)$ 之差
- 对 $d(k)$ 进行量化编码, 得到 $\tilde{d}(k)$





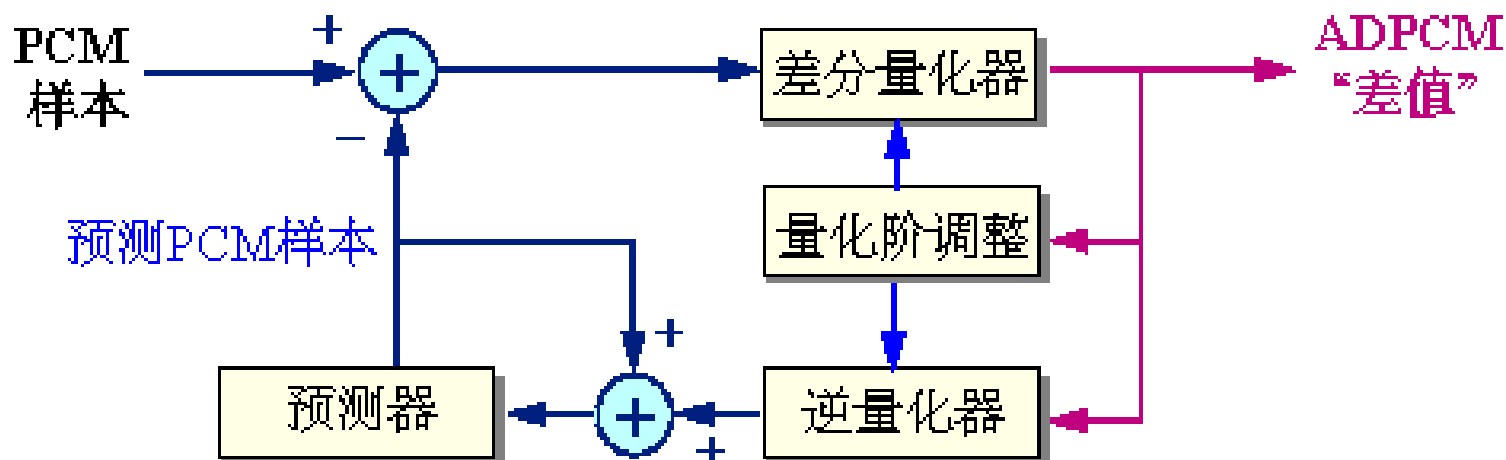
### 3.2.3 自适应差分脉冲编码调制

- 自适应差分脉冲编码调制 (Adaptive Differential Pulse Code Modulation, ADPCM)
  - 综合APCM的自适应特性和DPCM系统的差分特性，性能较好
  - 核心想法：
    - 利用自适应的思想改变量化阶的大小，即使用小的量化阶(step-size)去编码小的差值，使用大的量化阶去编码大的差值
    - 使用过去的样本值估算下一个输入样本的预测值，使实际样本值和预测值之间的差值总是最小

## 3.2.3 自适应差分脉冲编码调制

### ■ ADPCM编码框图

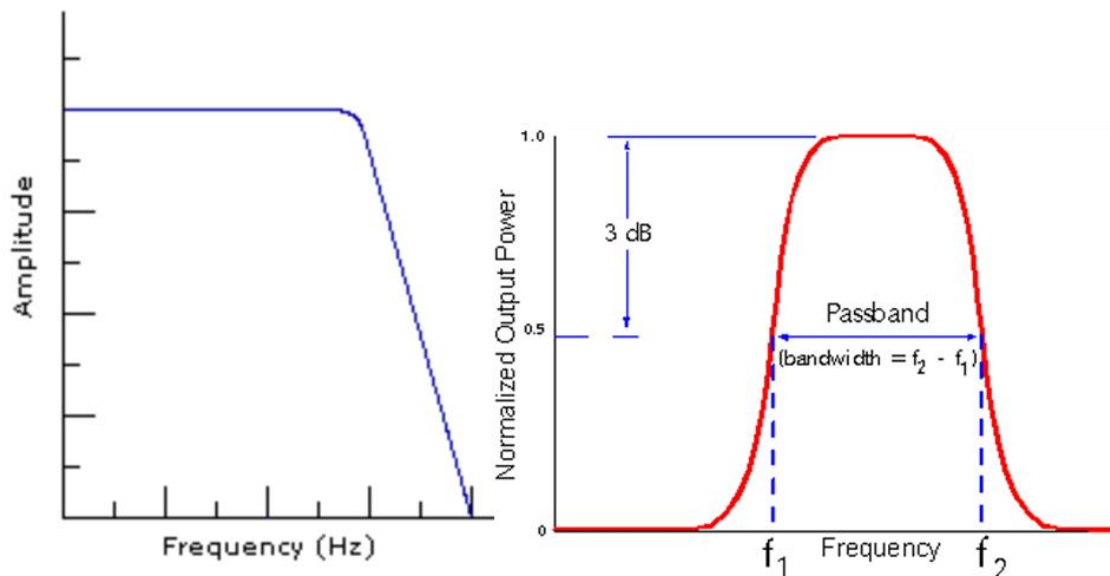
- 接收端的译码器使用与发送端相同的算法，利用传送来的信号来确定量化器和逆量化器中的量化阶大小，并且用它来预测下一个接收信号的预测值



### 3.2.3 G.722 SB-ADPCM编译码器

#### ■ 子带编码(sub-band coding, SBC)

- 用一组带通滤波器(band-pass filter, BPF)把输入声音信号的频带分成若干个连续的频段，每个频段称为子带。对每个子带中的声音信号采用单独的编码方案去编码



低通滤波器

带通滤波器

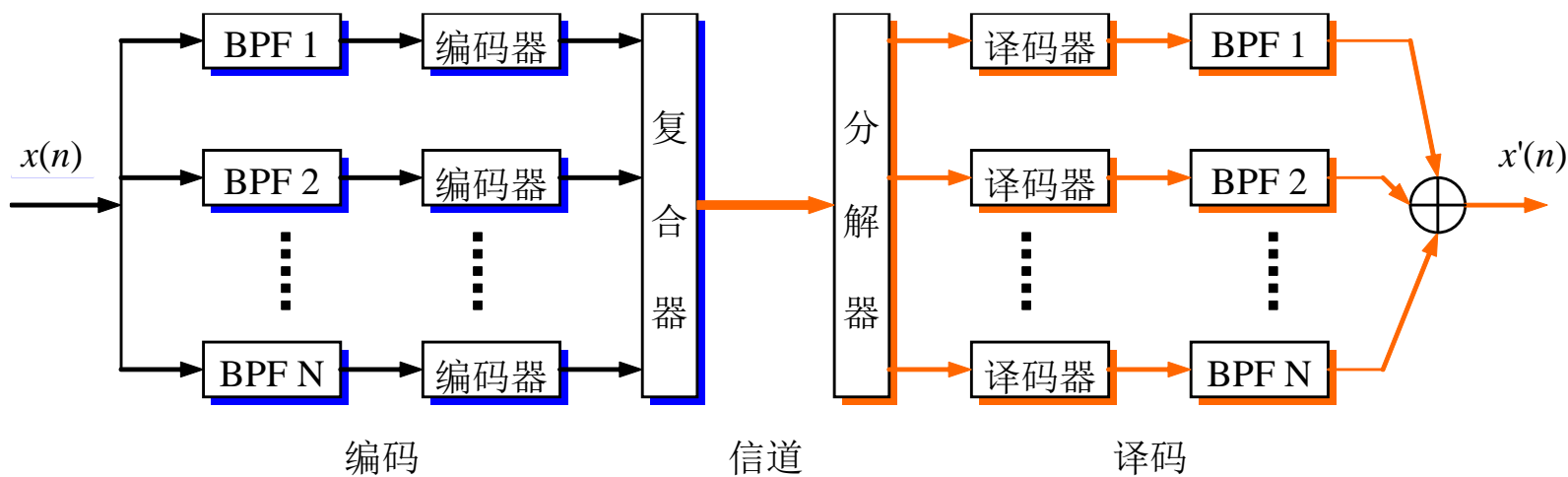


### 3.2.3 G.722 SB-ADPCM编译码器

- 在信道上传送时，将每个子带的代码复合在一起；在接收端译码时，将每个子带的代码单独译码，然后把它们组合起来，还原成原来的声音信号
- 好处：
  - 可对每个子带信号分别进行自适应控制，量化阶的大小可按照每个子带的能量电平加以调节。
  - 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。

### 3.2.3 G.722 SB-ADPCM编译码器

- 编码/译码器可以采用ADPCM, APCM或PCM



子带编码方块图



## 3.2.3 G.722 SB-ADPCM编译码器

### ■ 声音频带的分割

- 可用树型结构的式样进行划分
  - 等宽：首先把整个声音信号带宽分成两个相等带宽的子带：高频子带和低频子带, 然后重复对子带采用同样的划分方法。
  - 不等宽：把信号宽度分为不等带宽的子带, 如  $[0,500)$ ,  $[500,1000)$ ,  $[1000,2000)$ ,  $[2000,3000)$ 和  $[3000, 4000]$
- 为了能抵消混迭效应, 可采用正交镜像滤波器 (quadrature mirror filter, QMF)来划分频带



## 3.2.4 线性预测编码 (LPC)

### ■ 线性预测编码(Linear Predictive Coding, LPC)

- 将话音生成机理模型化为一个离散的、时变的、线性的递归滤波器
- 编码时使用线性预测分析话音波形产生声道激励和转移函数的参数，对声音波形的编码实际就转化为对这些参数的编码，这就可减少声音的数据量





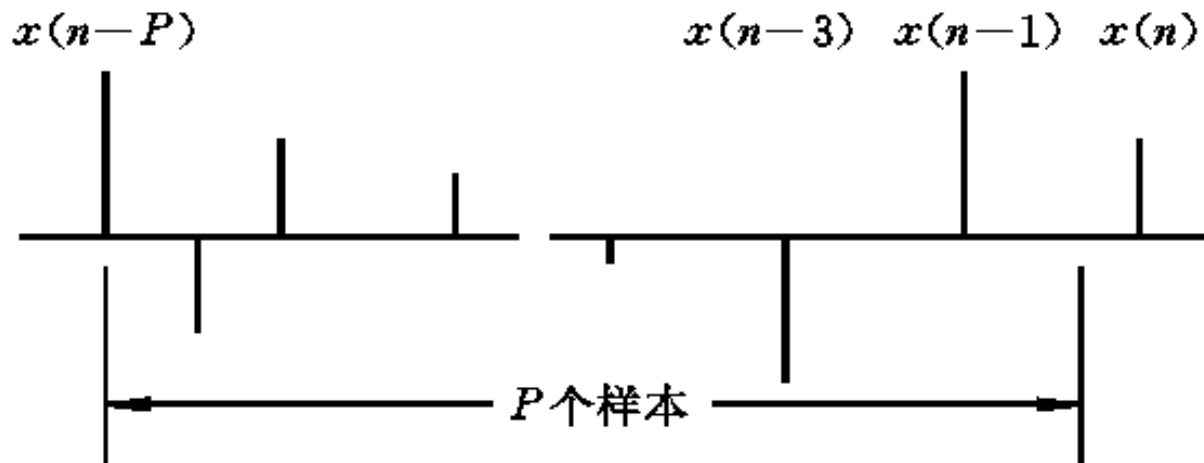
## 3.2.4 线性预测编码 (LPC)

- 译码时使用线性预测分析得到的参数，通过话音合成器重构话音。合成器实际上是一个离散的随时间变化的时变线性滤波器，它代表人的话音生成系统模型
- 时变线性滤波器
  - 分析话音波形时，当作预测器使用
  - 合成话音波形时，当作生成模型使用

## 3.2.4 线性预测编码(LPC)的概念

### ■ 线性预测器

- 使用过去的 $P$ 个样本值来预测现时刻的采样值 $x(n)$



LPC预测

## 3.2.4 线性预测编码(LPC)的概念

- 预测值用过去 $P$ 个样本值的线性组合表示为

$$\begin{aligned}x_{pre}(n) &= -[a_1x(n-1) + a_2x(n-2) + \cdots + a_px(n-p)] \\ &= -\sum_{i=1}^p a_i x(n-i)\end{aligned}$$

为方便起见，式中采用了负号  
残差误差即线性预测误差为一个线性差分方程，

$$e(n) = x(n) - x_{pre}(n) = \sum_{i=0}^p a_i x(n-i)$$

在给定的时间范围里，如 $[n_0, n_1]$ ，使 $e(n)$ 的平方和即

$$\beta = \sum_{n=n_0}^{n_1} [e(n)]^2$$

为最小，这样可使预测得到的样本值更精确。



## 3.2.4 线性预测编码(LPC)的概念

### ➤ 预测系数

- 通过求解偏微分方程，可找到系数 $a_i$ 的值
- 如果把发音器官等效成滤波器，这些系数 $a_i$ 可以理解成滤波器的系数
- 在接收端重构的话音不再具体复现真实话音的波形，而是合成的声音



### 3. 2. 4 GSM编译码器简介

#### ■ 全球数字移动通信系统 (Global System for Mobile communications, GSM)编译码器简介

- GSM算法是1992年柏林技术大学(Technical University Of Berlin)根据GSM协议开发的, 这个协议是欧洲最流行的数字蜂窝电话通信协议。
- GSM算法是一种使用较普遍的波形声音压缩算法, 使用了LPC码。



## 3.2.4 GSM编译码器简介(续)

### ■ GSM的性能

- GSM的输入：数据分成帧(frame)，一帧(20毫秒)由带符号的160个样本组成，每个样本为13位或16位的线性PCM(linear PCM)码
  - 使用的采样频率为8 kHz时，如果每个样本为16位，那么未压缩的话音数据率为
- GSM的输出：一帧( $160 \times 16$ 位)的数据压缩成260位的GSM帧，相当于13 kb/s。由于260位不是8位的整数倍，因此编码器输出的GSM帧为264位的线性PCM码
  - 使用GSM压缩后的数据率为：
- GSM的压缩比：



# 总结

---

- 声音、声音数字化
- 声音的编码
  - PCM
  - DM, ADM
  - APCM, DPCM, ADPCM
  - LPC

# 常见音频编码格式

vivo 智能手机Hi-Fi  
品质，邀您畅听

标准品质 (MP3/128kbps)

💎 HQ高品质 (MP3/320kbps)

💎 SQ无损品质 (APE)

💎 SQ无损品质 (FLAC)

🔴 限量 5.1声道 (FLAC)

## ■ 常见音频编码格式



- MP3 (MPEG-1 and/or MPEG-2 Audio Layer III) 有损压缩



- AAC(Advanced Audio Coding):包含在MPEG2, MPEG4中；有损压缩；相同速率下音质比MP3更好。



- APE(Monkey's Audio):无损压缩，压缩比比FLAC大，解码更复杂



- FLAC(Free Lossless Audio Codec)：无损压缩，50%-60%



END

---

## 第3章 数字声音编码

