



Research Institute for Future Media Computing    Institute of Computer Vision  
未来媒体技术与研究所    计算机视觉研究所



# 多媒体系统导论

## Fundamentals of Multimedia System

授课教师：朱映映教授

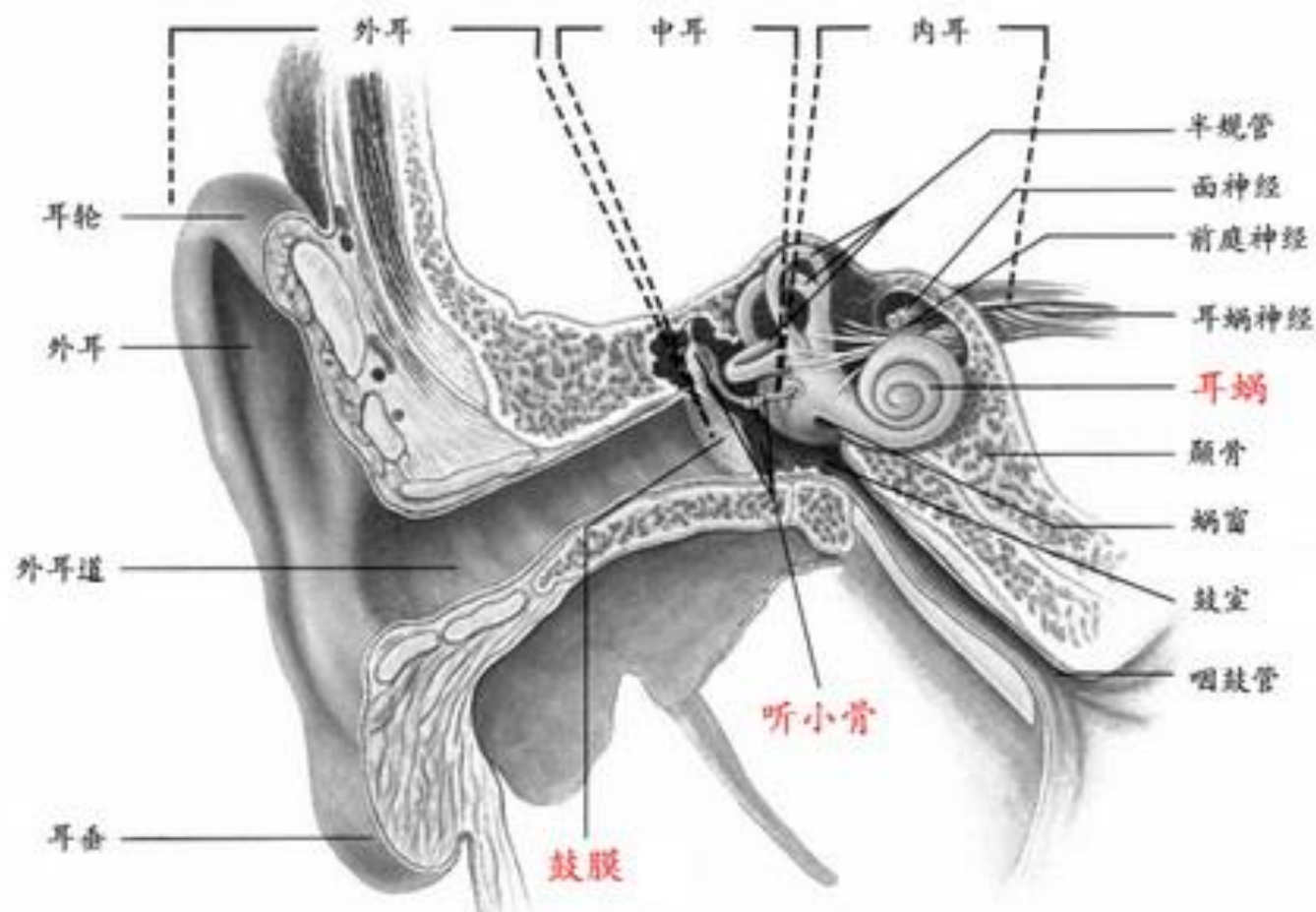
Email: zhuyy@szu.edu.cn

# 第四讲

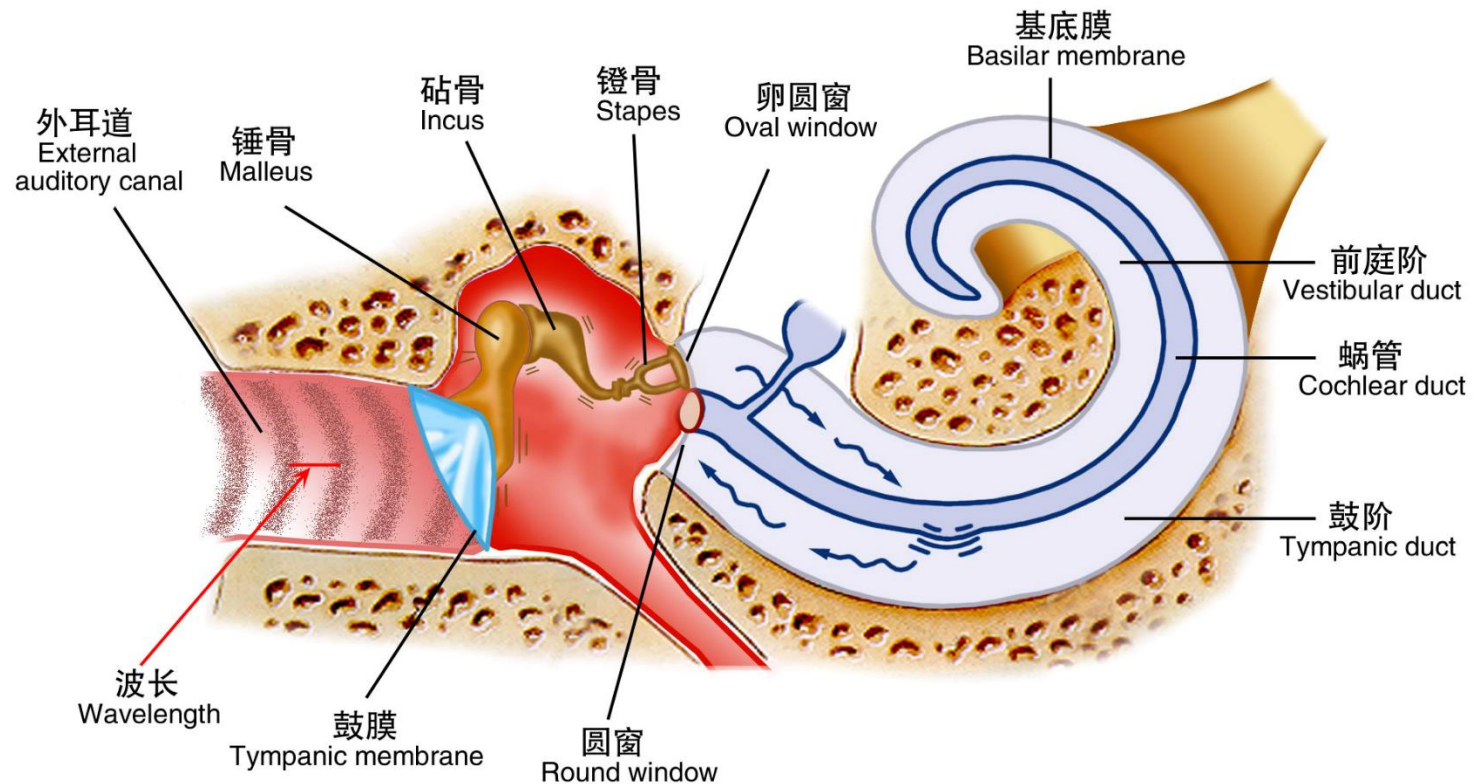
## Basics of digital Audio

### 第6章

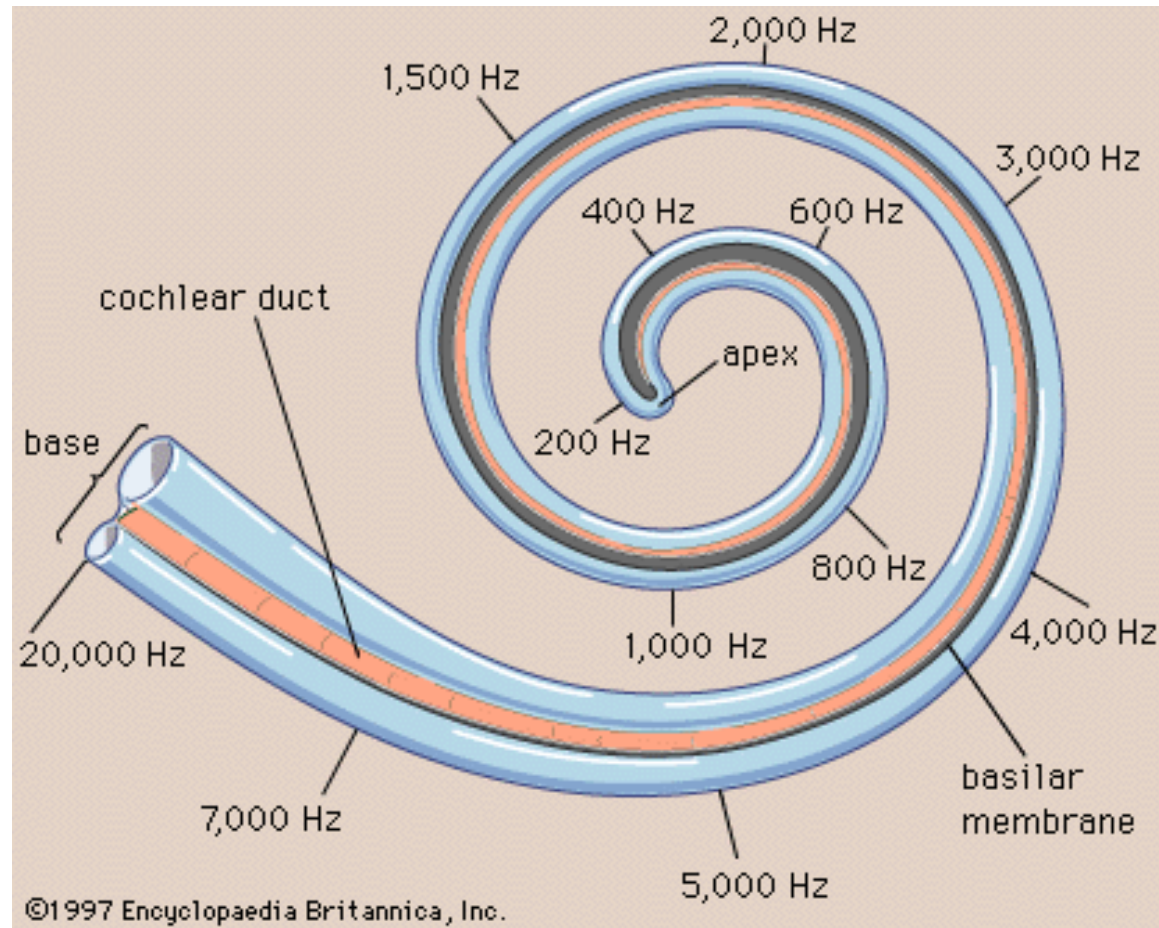
# Human Ear



# Cochlea and Its Function



# Cochlea and Its Function



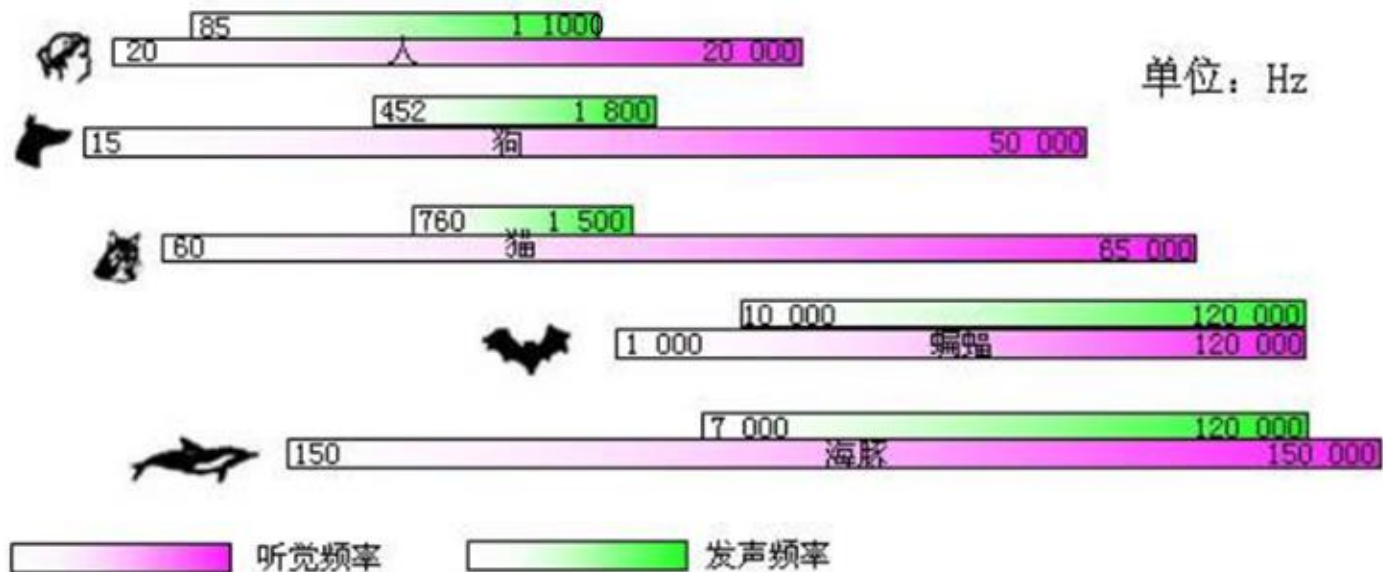
# How Sound is Perceived

- ◆ The cochlea (耳蜗) detects sound
  - An organ in our inner ears
  - The cochlea is joined to the eardrum (鼓膜) by three tiny bones
- ◆ The cochlea consists of a spiral (螺旋形) of tissue filled with liquid and thousands of tiny hairs
  - Each hair is connected to a nerve which feeds into the auditory nerve bundle going to the brain
  - The longer hairs resonate (共振) with lower frequency sounds, and the shorter hairs with higher frequencies
- ◆ The cochlea serves to transform the air pressure signal experienced by the eardrum into frequency information
  - Can be interpreted by the brain as sound

# Basics of Sound

## ◆ What is Sound?

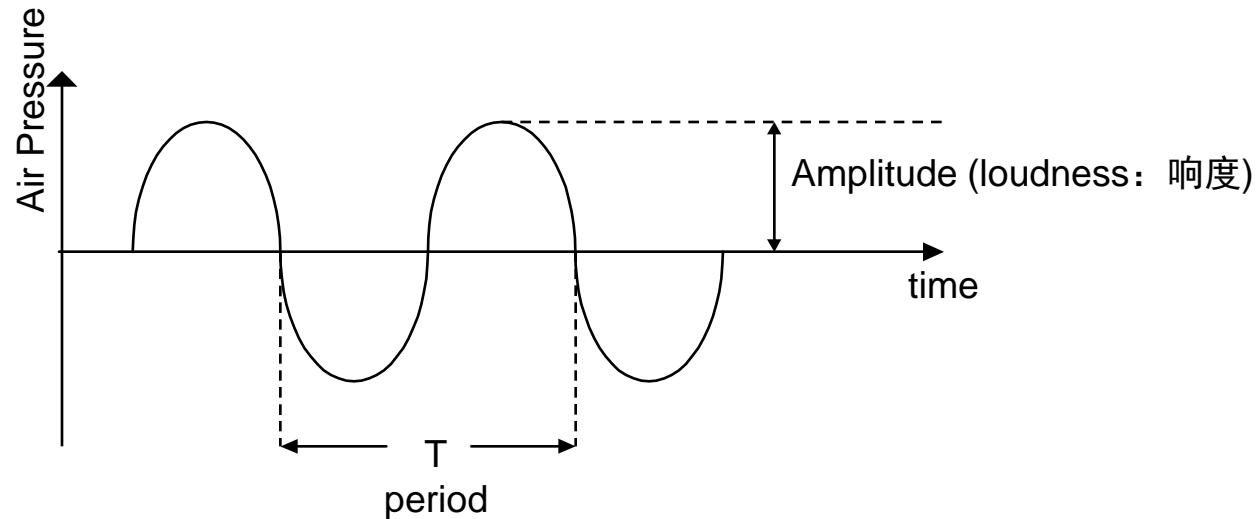
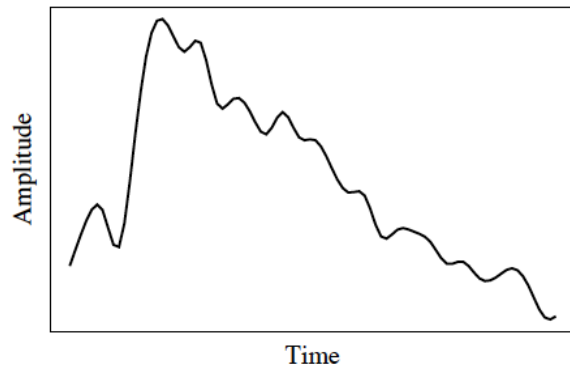
- Sound (声音) is a pressure wave that moves through a compressible (可压缩的) medium
- Frequency ranges of human and animals.



# Basics of Sound

## ◆ Characteristics of sound

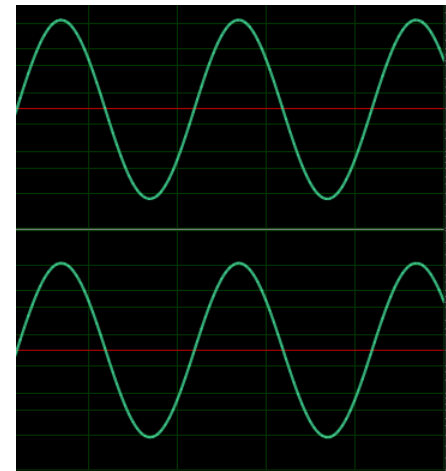
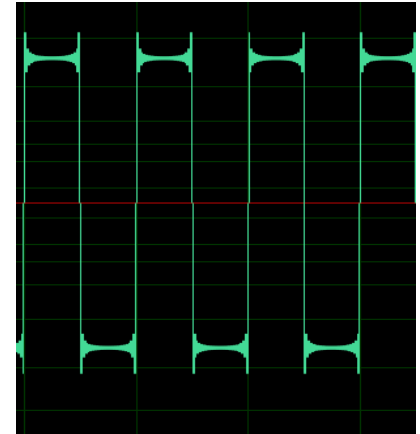
- Amplitude (振幅)
  - Measure the pressure at a location using voltage levels (电平)
- Frequency
  - The number of periodic vibrations (震动) within a second





# Basics of Audio

- ◆ Audio (音频) is the sounds human can hear
- ◆ Frequency response
  - Refers to the range of frequencies
  - 20Hz-20kHz for human hearing
  - Human are more sensitive to low frequencies
- Two samples
  - Square/Rectangular wave 1kHz
  - Sine/Sinusoid wave 1kHz



# Magnitude Range of Audio

## ◆ Magnitude（量级） levels

- The levels of audio are described in terms of decibels, as a ratio to the quietest sound we are capable of hearing
- Humans perceive sounds over the entire range of 120dB, the upper limit of which will be painful to humans

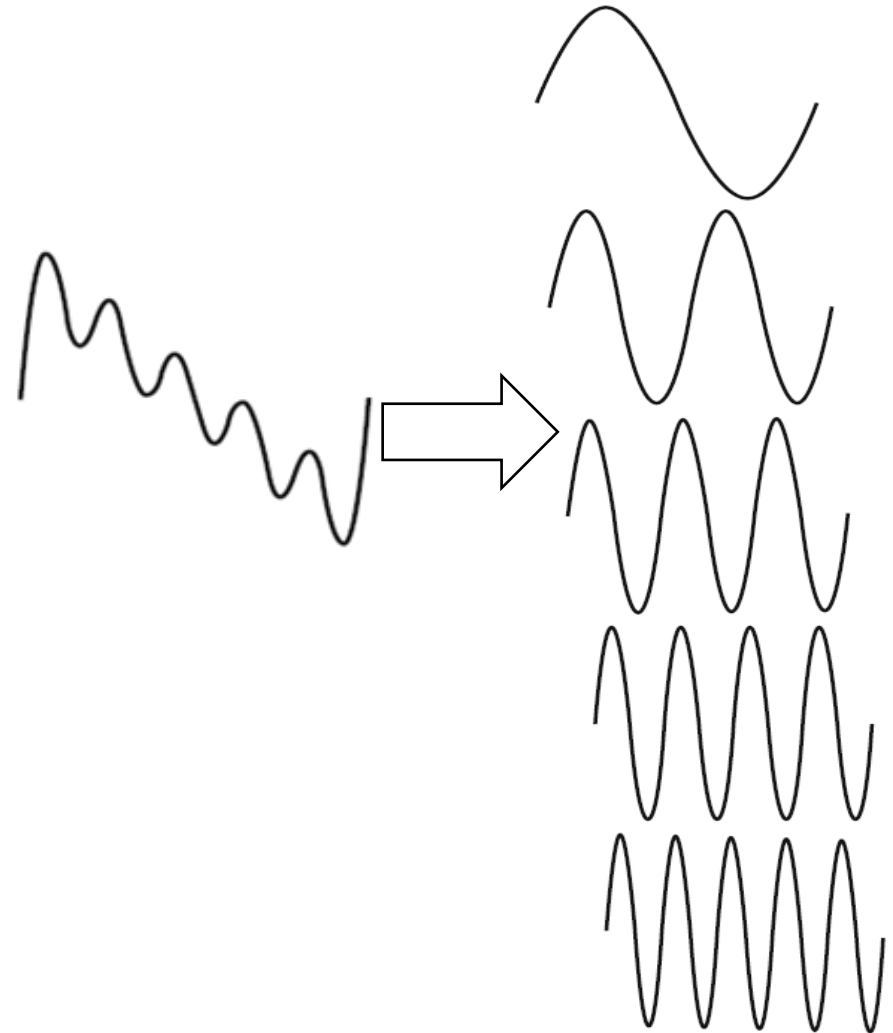
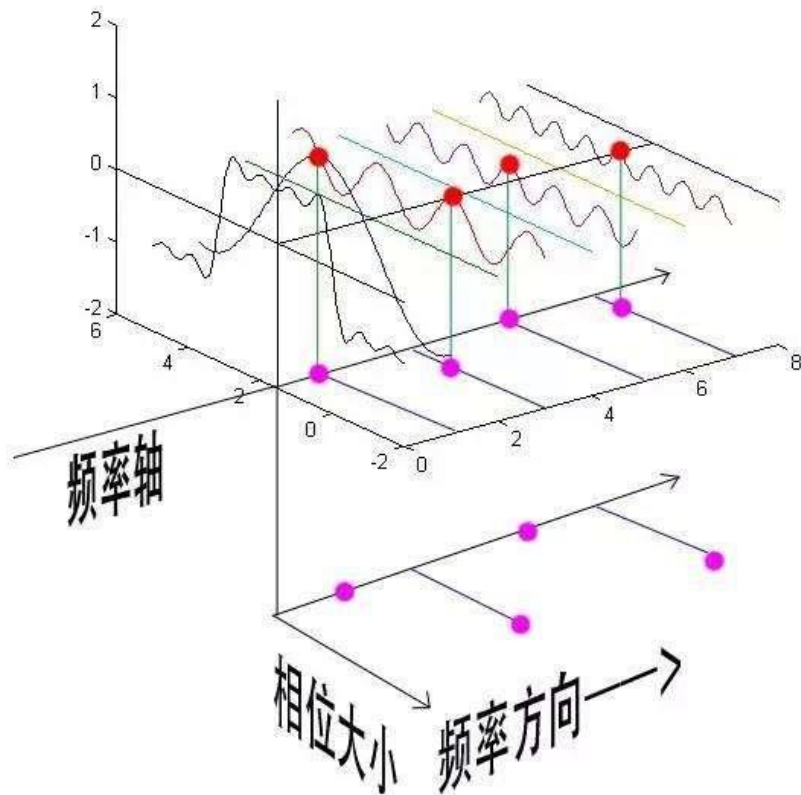
音量	相当于
0 dB	听力的阈值
60 dB	日常对话
70 dB	嘈杂的餐馆
80 dB	嘈杂的交通，工厂噪音，吸尘器
90 dB	地铁列车，割草机
100 dB	电锯，冲击钻
120 dB	不舒服的阈值（摇滚音乐会，打雷）
140 dB	痛苦的阈值
160 dB	伤及鼓膜

# Basics of Speech

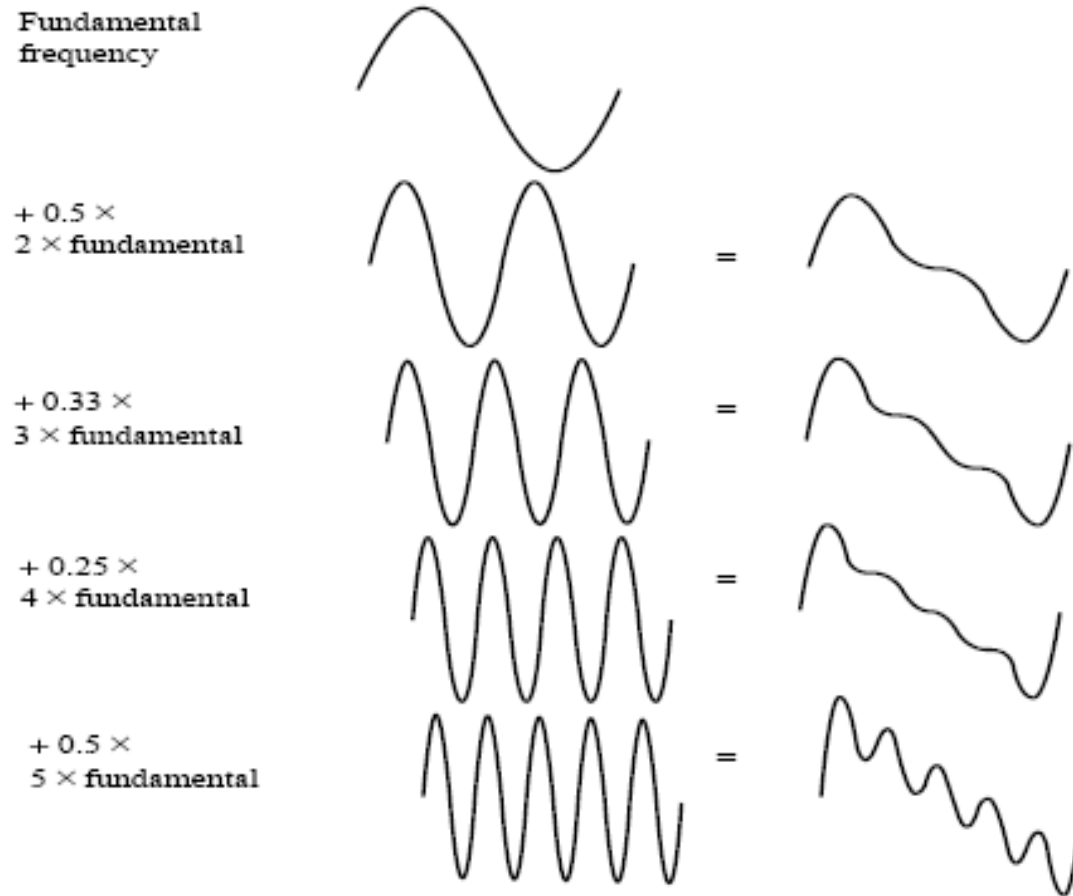
- ◆ Speech (语音) is the sounds human can utter (说出)
  - The human ear is most sensitive in the range from 600HZ to 6,000HZ
  - 300-3400Hz for human speech
  - A human adjusts himself/herself very efficiently to different speakers and their speech habits
- ◆ Speech analysis
  - Speaker recognition
    - Recognize a speaker for his identification and verification
  - Speech recognition
    - Analyze what have been said and how a certain statement was said
  - Speaker dependent systems
    - Higher recognition
    - Brittle (脆弱的) in terms of speaker
  - Speaker independent systems
    - Lower recognition
    - Many speakers

# Building up a complex signal by superposing sinusoids

- ◆ Signals can be decomposed into a sum of sinusoids



# Building up a complex signal by superposing sinusoids



Fundamental Frequency,  $f = 100\text{Hz}$

# 声音信号数字化

## ◆ 1. 从模拟过渡到数字

- 大多数电信号是模拟信号，用模拟器件处理难度大，精度低，成本高
- 用数字表示模拟量：把模拟信号转变成数字信号，使用数字信号处理器(digital signal processor, DSP)
- 在数字域而不在模拟域中做信号处理的主要优点
  - 数字信号计算是精确运算，不受时间和环境变化影响
  - 复杂的数学运算容易实现
  - 可对数字运算部件进行编程

# 声音信号数字化(续)

## ◆ 2. 模拟信号与数字信号

- 模拟信号：时间和幅度上都是连续的信号
- 数字信号：时间和幅度都用离散值表示的信号
  - 在特定时刻对模拟信号进行测量叫做采样(sampling)，由这些特定时刻采样得到的信号称为离散时间信号
  - 采样得到的幅值是无穷多个实数值中的一个。如果把信号幅度取值的数目加以限定，这种由有限数目的数值组成的信号称为离散幅度信号
  - 【例】假设输入电压范围为 $0.0 \sim 0.7 \text{ V}$ ，其取值只限定为 $0, 0.1, 0.2, \dots, 0.7$ 共8个值
    - 如果采样得到的幅度值是 $0.123 \text{ V}$ ，它的取值就算作 $0.1 \text{ V}$
    - 如果采样得到的幅度值是 $0.26 \text{ V}$ ，它的取值就算作 $0.3 \text{ V}$

# 声音信号数字化(续)

## ◆ 3. 声音数字化方法

### - (1) 数字化的概念

- 声音进入计算机的第一步就是数字化
  - 模拟信号在时间和幅度都用离散值表示
  - 图1表示了声音数字化的概念

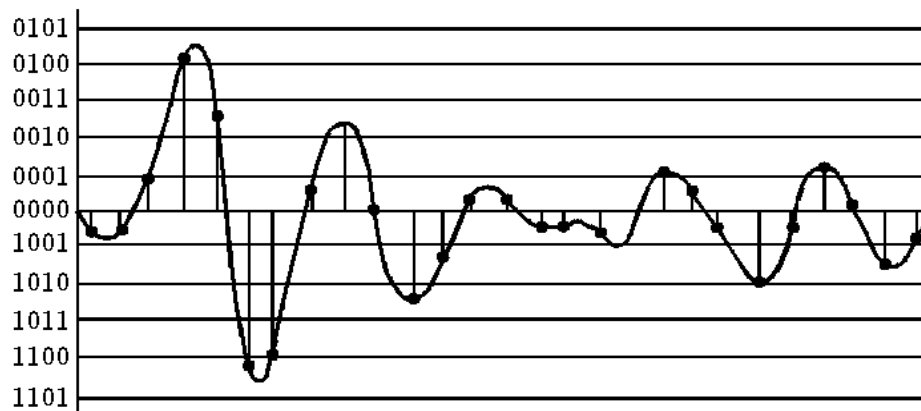


图1 声音的采样和量化



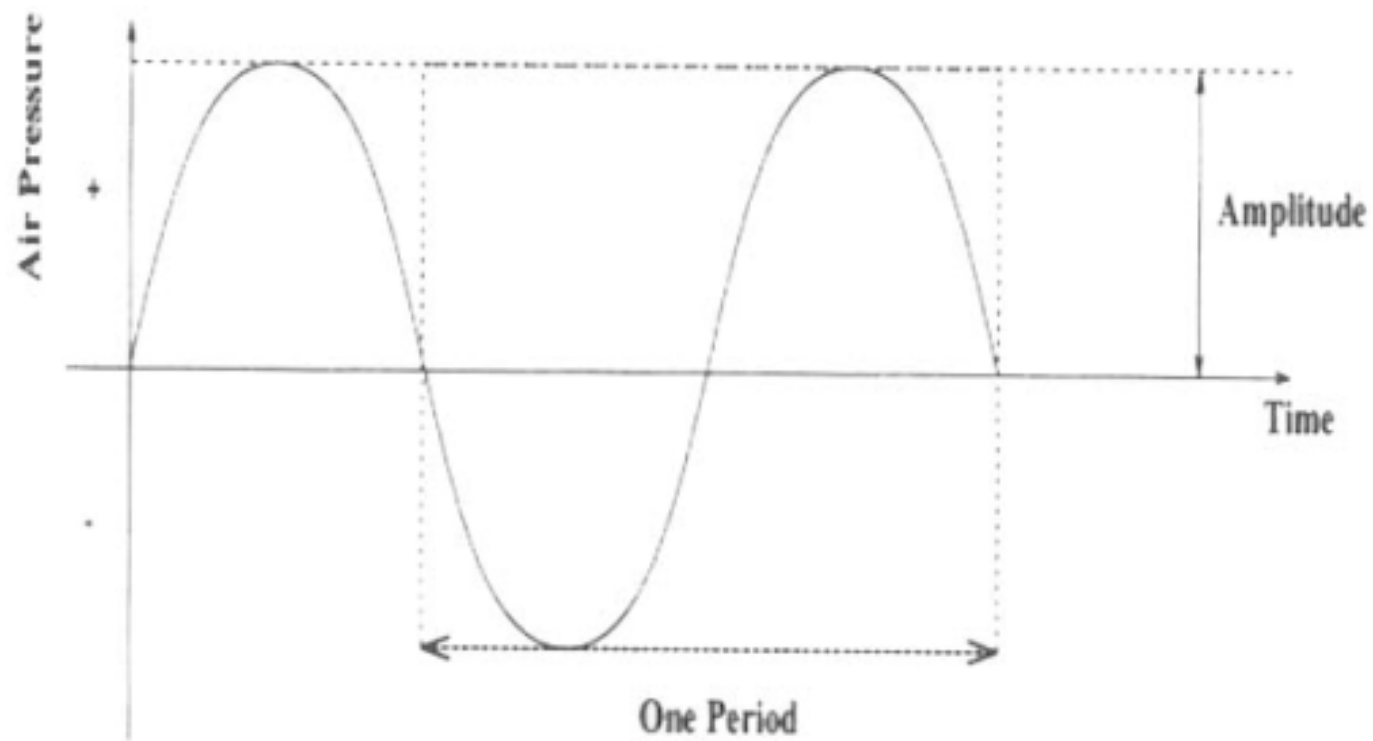


图2 声音的原始波形示例

# 采样

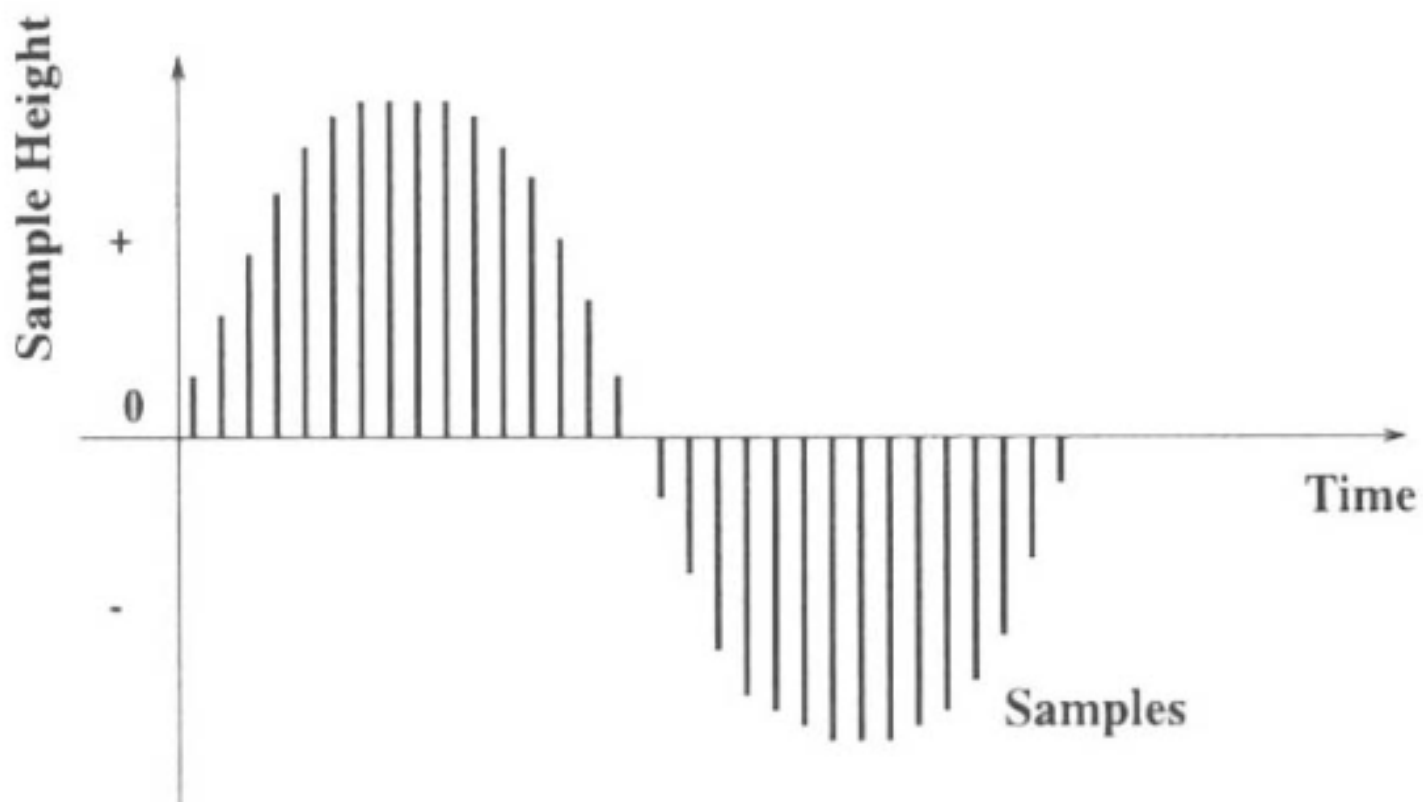


图3 声音的采样

## 量化

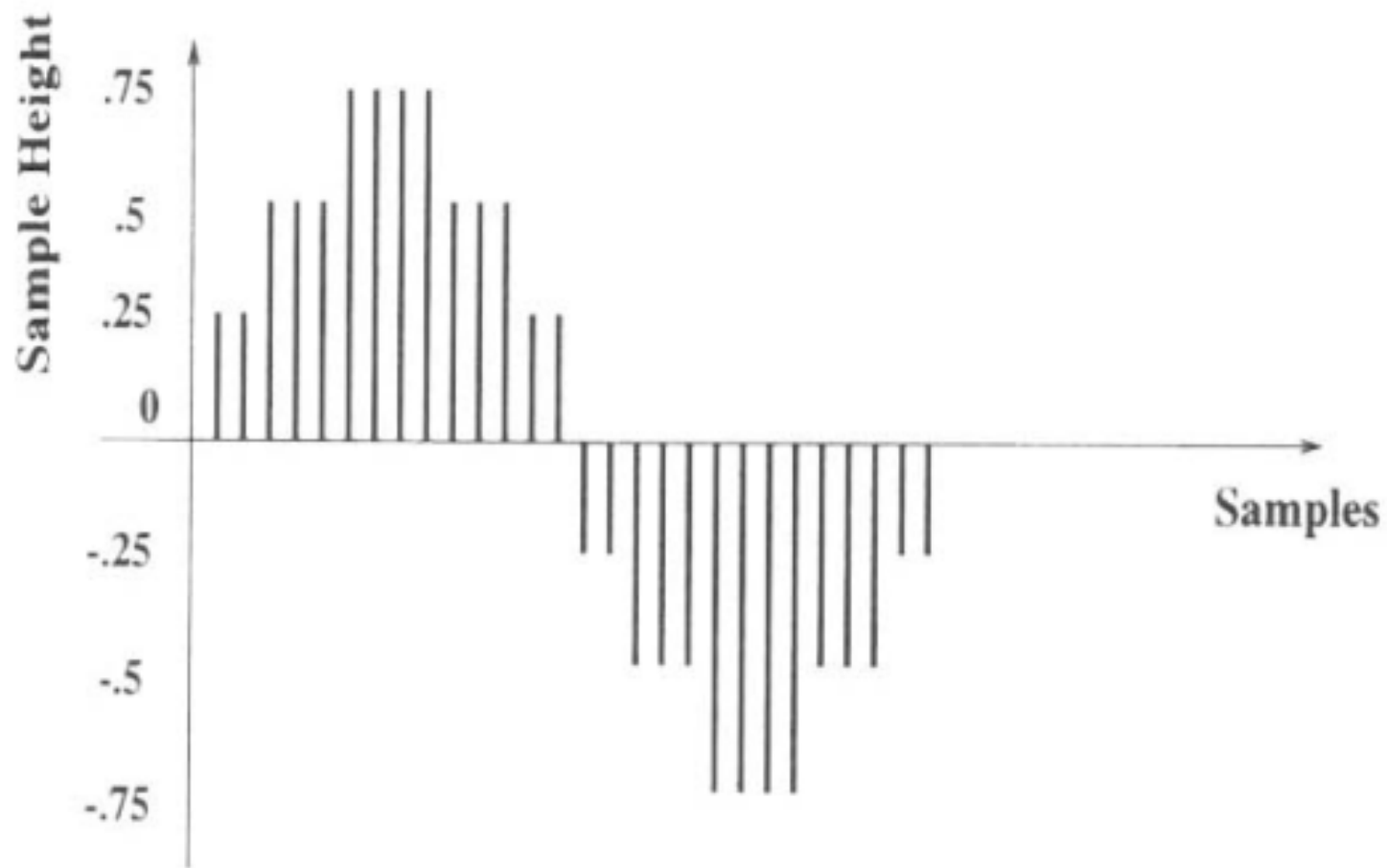


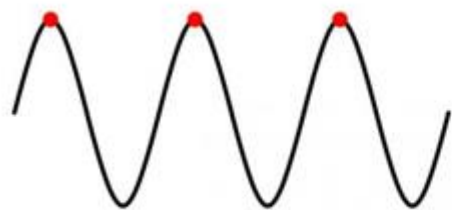
图4 声音的量化

# 声音信号数字化(续)

## (2) 奈奎斯特(Nyquist)定理

- 奈奎斯特 (Nyquist) 定理, 也称为奈奎斯特采样定理, 是数字信号处理和通信工程中的一项基本原理, 由美国物理学家哈维·奈奎斯特在20世纪早期提出。该定理描述了理想情况下, 对模拟信号进行数字化采样时所需的最低采样频率, 以便能准确无损地重建原始信号。
- 其核心内容: 若要从一个带限 (即信号频谱中最高频率有限) 的模拟信号中无失真地恢复原信号, 采样频率 ( $f_s$ ) 必须大于或等于信号最高频率成分 ( $f_N$ ) 的两倍。数学表达式为:  $f_s > 2 * f_N$
- 如果信号的采样率低于两倍奈奎斯特频率, 采样数据中就会出现虚假的低频成分。这种现象便称为频率混叠 (Frequency Aliasing), 也称为假频。

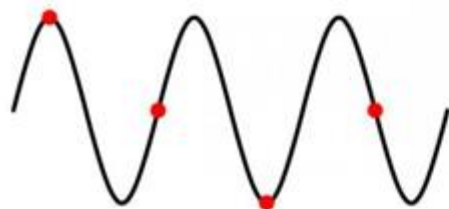
# 声音信号数字化(续)



**A**  
→  
Sampled at  $f$



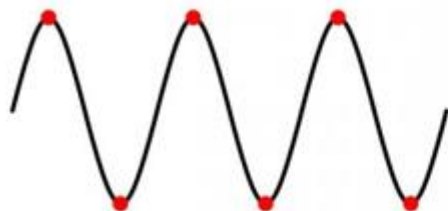
$$f_s = f_N$$



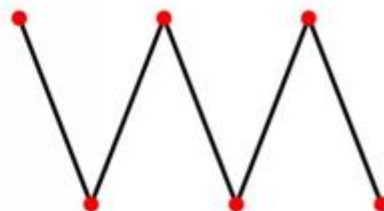
**C**  
→  
Sampled at  $4f/3$



$$f_s = \frac{4}{3} * f_N$$



**B**  
→  
Sampled at  $2f$



$$f_s = 2 * f_N$$

# 声音信号数字化(续)

## (3) 采样精度

- 度量声音波形幅度的精确程度，通常用存储每个声音样本的位数( $n$ )来表示
  - 例如，每个声音样本用16位(2字节)表示，测得的声音样本值是在 $[0 \sim 65535]$ 范围里的数，采样精度是 $1/65536$
- 精度是度量模拟信号的最小单位，称为量化阶(quantization step size)
  - 如将 $0 \sim 1$  V的电压用 $n=256$ 个数表示，其量化阶等于 $1/256$  V。
- 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
  - 位数越少，声音质量越低，所需存储空间也越少

# 声音信号数字化(续)

## (4) 信号噪声比与信号量化噪声比

- 信号的能量与噪声能量的比例称为**信号噪声比**，简称为**信噪比**(signal to noise ratio, SNR)，用下式计：

$$SNR = 10 \log_{10} \frac{V_{\text{signal}}^2}{V_{\text{noise}}^2} = 20 \log_{10} \frac{V_{\text{signal}}}{V_{\text{noise}}}$$

- 信噪比是信号质量的衡量标准之一，单位为分贝(dB)。
- 除了原始的模拟信号中存在的噪声外，离散化会带来其他一些误差。
- 量化噪声（量化误差）：某个采样时间点的模拟信号的采样值与它最接近的数字数值之间的差值。
- **信号量化噪声比SQNR**: 衡量量化质量

$$\begin{aligned} SQNR &= 20 \log_{10} \frac{V_{\text{signal}}}{V_{\text{quan\_noise}}} = 20 \log_{10} \frac{2^{N-1}}{\frac{1}{2}} \\ &= 20 \times N \times \log 2 = 6.02N(\text{dB}) \end{aligned}$$

假设每个采样点的量化精度为N位，其中， $V_{\text{signal}}$ 取峰值为 $2^{N-1}$ ，同时分母 $V_{\text{quan\_signal}}$ 取最大值 $1/2$ 。

## 声音信号数字化(续)

- ◆ N越大, SQNR越大, 这说明对模拟信号的逼近越精确, 系统能够提供的音质越好。
- ◆ 例, 假设信号电压 $V_{\text{signal}}=0.7\text{V}$ , 采样精度用16bit表示, 求信号量化噪声比。

$$\text{SQNR}=6.02*16=96(\text{dB})$$



# 声音信号数字化(续)

## (5) 均匀量化(也称线性量化)

- 采用相等的量化间隔对采样信号进行量化, 见图5
- 量化误差或量化噪声: 样本值  $Y$  和原始值  $X$  之差  $E = Y - X$

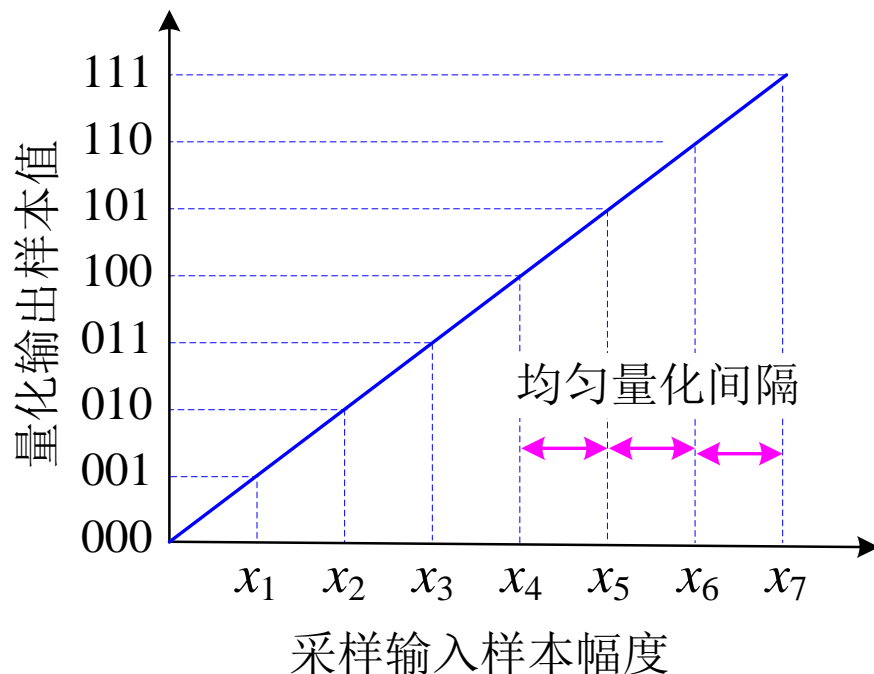


图5 均匀量化

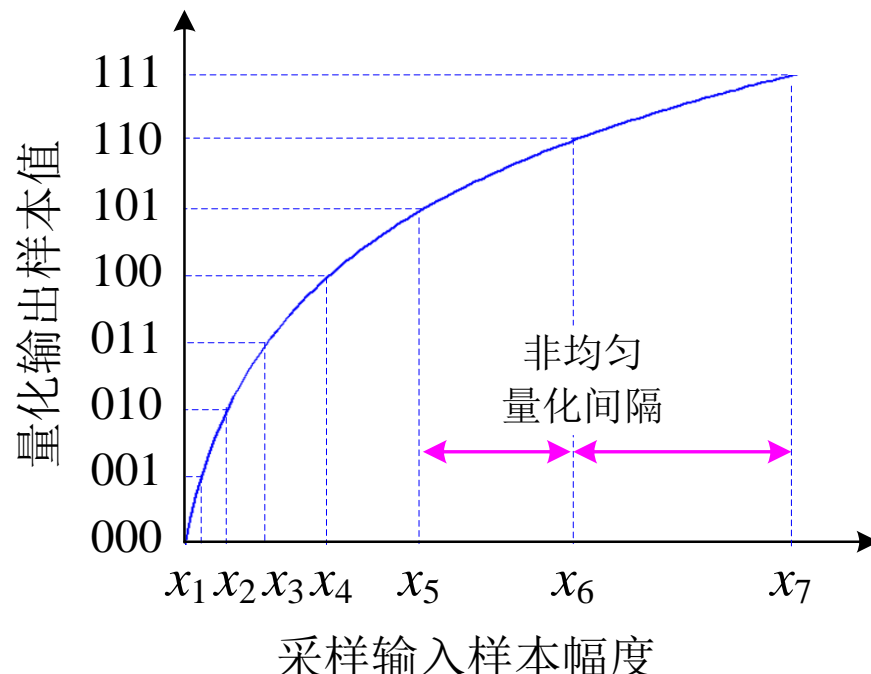


图 6 非均匀量化

## (6) 非均匀量化

- 基本想法：对幅度大的信号采用大的量化间隔，幅度小的信号采用小的量化间隔，如图 6所示。
- 可在满足精度要求的情况下用较少的位数来表示。
- 定义了两种算法： $\mu$ 律( $\mu$ -law)编码(图7)和A律(A-law)编码(图8)
- 这两种算法与人的听觉感知特性一致, 其中的对数函数对于振幅小的声音更敏感。

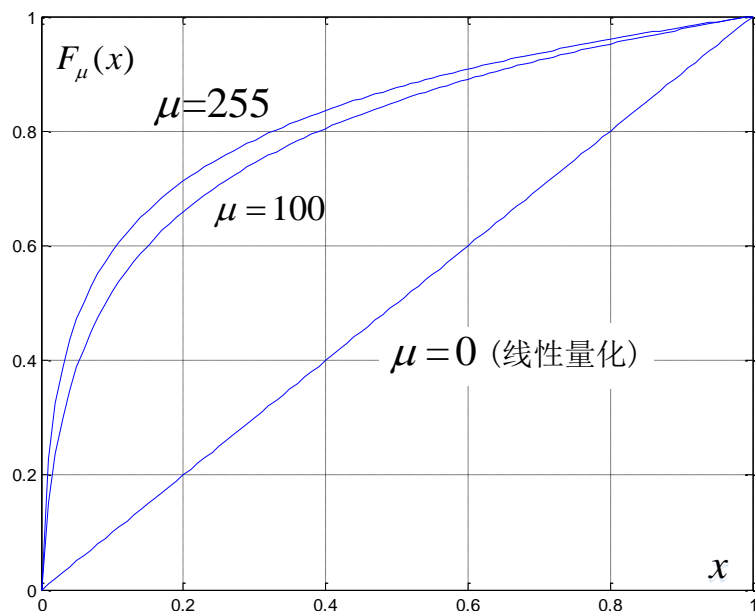


图7  $\mu$ 律编码特性

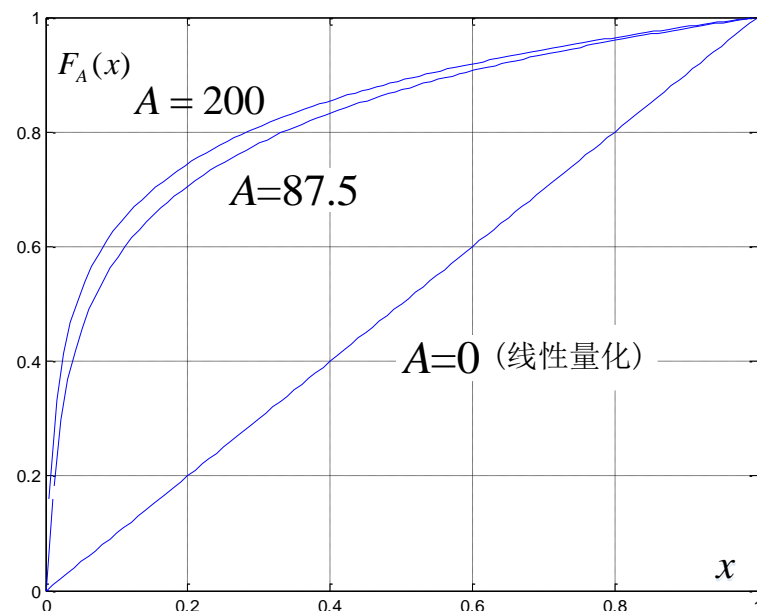


图8 A律编码特性

**(1)  $\mu$ 律压扩按下面的式子确定量化输入和输出的关系,**

$$F_{\mu}(x) = \operatorname{sgn}(x) \frac{\ln(1 + \mu |x|)}{\ln(1 + \mu)} \quad (4-2)$$

式中,  $x$  为输入信号幅度, 规格化成  $-1 \leq x \leq 1$ ; 函数  $\operatorname{sgn}(x)$  的值为 1, 符号同  $x$  的正负号;  $\mu$  为确定压缩量的参数, 它反映最大量化间隔和最小量化间隔之比, 取  $100 \leq \mu \leq 500$ 。

由于  $\mu$  律压扩的输入和输出关系是对数关系, 如图7所示, 所以这种编码又称为对数PCM。具体计算时, 用  $\mu = 255$ , 把对数曲线变成 8 条折线以简化计算过程

**(2) A 律压扩按下面的式子确定量化输入和输出的关系:**

$$\begin{aligned} F_A(x) &= \operatorname{sgn}(x) \frac{A|x|}{1 + \ln A}, & 0 \leq |x| \leq 1/A \\ F_A(x) &= \operatorname{sgn}(x) \frac{1 + \ln(A|x|)}{1 + \ln A}, & 1/A < |x| \leq 1 \end{aligned} \quad (4-3)$$

式中,  $x$  为输入信号幅度, 规格化成  $-1 \leq x \leq 1$ ;  $\operatorname{sgn}(x)$  为  $x$  的极性;  $A$  为确定压缩量的参数, 反映最大量化间隔和最小量化间隔之比。

A 律压扩的前一部分是线性的, 其余部分与  $\mu$  律压扩类似, 如图8 所示。具体计算时,  $A = 87.56$ , 为简化计算, 同样可把对数曲线部分变成折线。

- 编码：首先对模拟信号进行对数转换（电信业称为“压缩”），然后对其进行采样，再通过模/数转换器将采样结果数字化。模数转换器对“压缩”后的信号做均匀量化。
- 解码：首先使用数/模转换器，得到模拟信号后，通过一个“扩展”电路做逆对数转换。这个过程称为“压缩扩展”。
- 压缩扩展技术体现了一种非均匀量化的思想。

# 声音信号数字化(续)

## ◆ 4. 声音质量与数据率

- 根据声音频带，声音质量分成5个等级，由低到高
  - 电话(telephone)
  - 调幅(amplitude modulation, AM)广播声音
  - 调频(frequency modulation, FM)广播声音
  - 激光唱盘(CD-Audio) 声音
  - 数字录音带(digital audio tape, DAT) 声音

表 4-1 声音质量和数据率

质量	采样速率 (kHz)	样本精度 (bit/s)	单道声/ 立体声	数据率(kb/s) (未压缩)	频率范围
电话*	8	8	单道声	64.0	200~3 400 Hz
AM	11.025	8	单道声	88.2	20~15 000Hz
FM	22.050	16	立体声	705.6	50~7 000Hz
CD	44.1	16	立体声	1411.2	20~20 000 Hz
DAT	48	16	立体声	1536.0	20~20 000 Hz
* 电话使用 $\mu$ 律编码，动态范围为 13 位而不是 8 位					

此表中k=1024, 采样速率k 一般取1000

# 声音信号数字化(续)

## ◆ 5. 声音质量的MOS评分标准

- 衡量声音质量方法
  - 声音信号的带宽
  - 客观质量度量：主要用信噪比(SNR)
  - 主观质量度量：人的感觉更具决定意义
    - 主观平均分：MOS= (mean opinion score)
- 声音主观质量度量标准是5分制

表 4-2 声音质量 MOS 评分标准

分数	质量级别	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)

# 语音编码介绍

## - 1. 语音编码(speech coding)

- 针对语音数字数据的压缩技术，主要用于语音通信

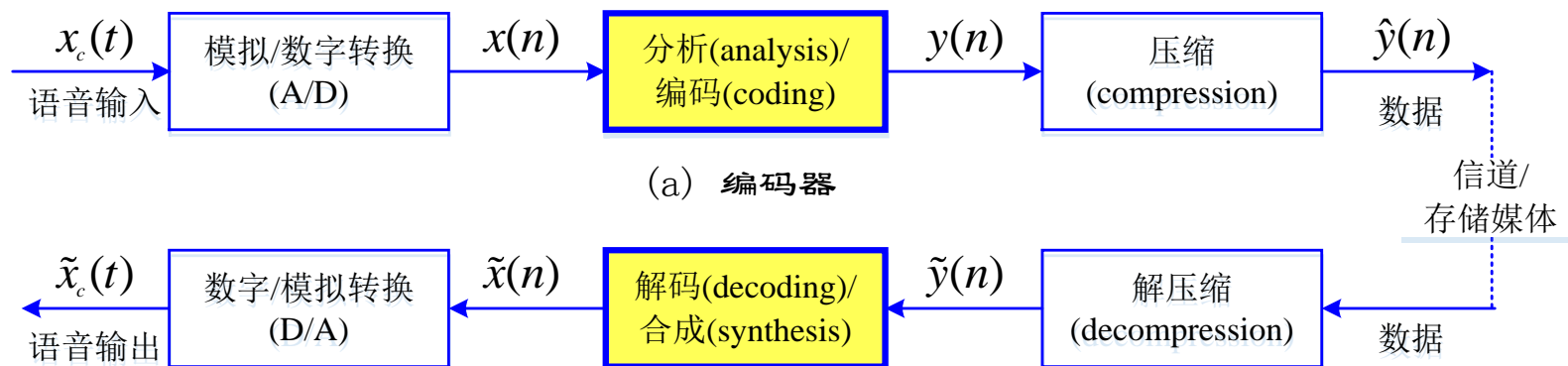


图9 语音编码概念

- 输入语音 经过A/D转换(采样和量化)后，通过分析和编码转换成代表语音的数字信号，解码过程与编码相反

# 语音编码介绍(续)

## • 编码方法：三种类型

- 波形编码(waveform coding)
- 参数编码(parametric coding)
- 混合编码(hybrid coding)

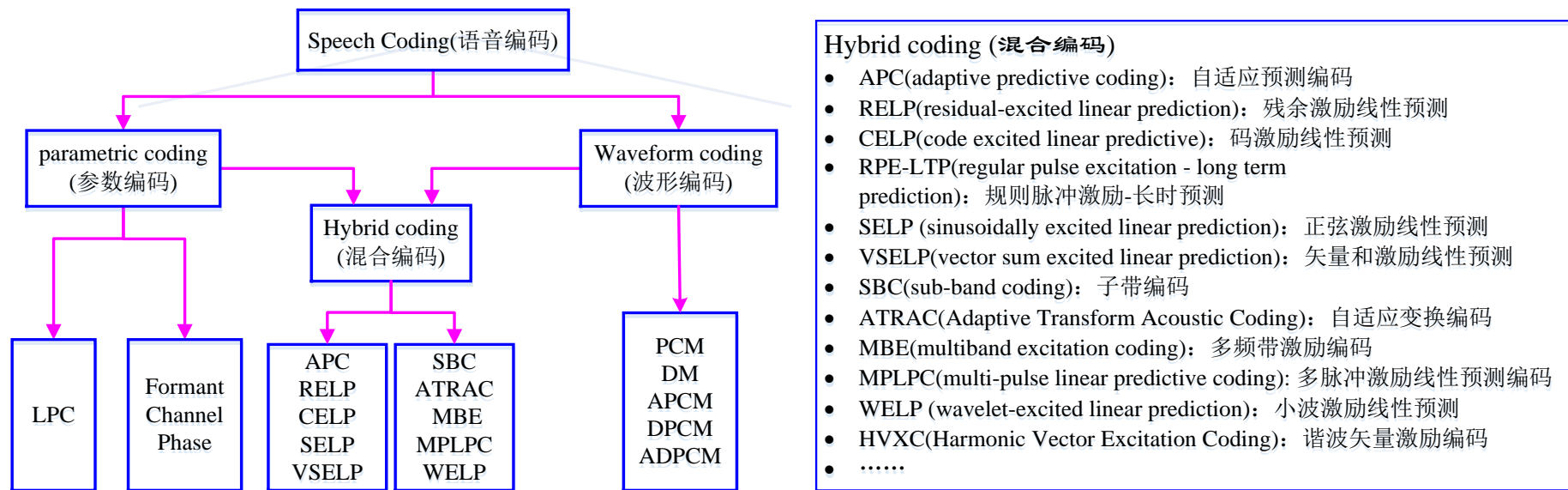


图10 语音编码方法类型



# 语音编码介绍(续)

## - 2. 波形编码

- 用数字形式精确地表示模拟信号波形的编码方法，不考虑语音产生和感知特性
- 编码器输出的数据速率为9.6~64 kbps或更高(取决于采样速率和量化精度)
- 算法包括
  - PCM(pulse code modulation): 脉冲编码调制
  - DM( $\Delta$ )(delta modulation, DM): 增量调制
  - APCM(adaptive pulse code modulation): 自适应脉冲编码调制
  - DPCM(differential pulse code modulation): 差分脉冲编码调制
  - ADPCM(adaptive difference pulse code modulation): 自适应差分脉冲编码调制

# 语音编码介绍(续)

## - 3. 参数编码

- 利用发音器官生成语音信号的模型，对从语音信号中抽出的语音特征参数(如发音模型、有声/无声、音量大小、音调)进行编码的方法
- 解码器根据模型参数重构语音信号
- 编码器输出的数据速率约为2~4.8 kbps
- 算法包括
  - LPC([Linear Predictive Coding](#)): 线性预测编码
  - Channel coding: 信道编码
  - Phase coding: 相位编码
  - Formant coding: 共振峰编码

# 语音编码介绍(续)

## - 4. 混合编码

- 综合使用波形编码和声源编码技术，组合波形特性和语音特征参数的编码方法
- 既有波形编码语音质量高的优点，又有参数编码数据速率低的优点
- 主要用在语音质量要求高的移动通信系统
- 混合编码器输出的数据速率为4.0 ~ 16 kbps
- 算法包括
  - APC(adaptive predictive coding): 自适应预测编码
  - RELP(residual-excited linear prediction): 残余激励线性预测
  - CELP(code excited linear predictive): 码激励线性预测
  - SBC(sub-band coding): 子带编码

# ■ 波形编码

## ◆ 基本思想

- 使重构的语音信号的波形与原始信号的波形尽量接近

## ◆ 语音质量

- 是语音编码质量最好的编码方法

## ◆ 编码方法

- PCM, DM, ADM, APCM, DPCM, ADPCM和SB-ADPCM

# ◆ 1. PCM编码(G.711)

- 脉冲编码调制(pulse code modulation, PCM)
  - 概念最简单、理论最完善的编码系统
  - 最早研制成功、使用最为广泛的编码系统
  - 数据量最大的编码系统
  - 1972年ITU将它作为G.711声音(audio)编码标准

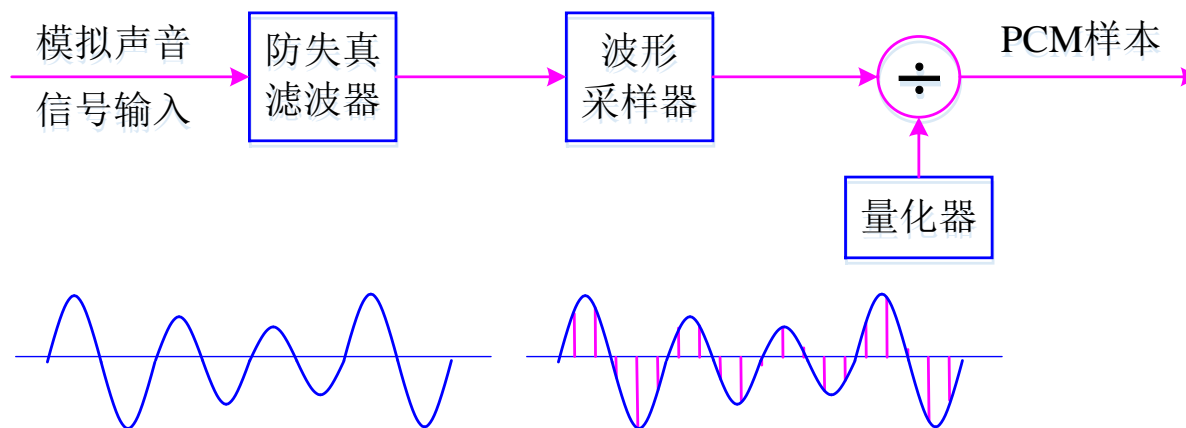


图11 PCM编码框图

- 防失真滤波器：低通滤波器，滤除声音频带外的信号

# ■ 波形编码(续)

- PCM编码用在时分多路复用的语音通信中
  - 时分多路复用(TDM=time-division multiplexing)
    - 在同一条通信线路上使用不同时段“同时”传送多个独立信号
    - 核心思想：将时间分成等间隔的时段，为每个用户指定一个时间间隔，使许多用户同时使用一条传输线路
  - 语音信号的采样速率  $f = 8000 \text{ Hz/s}$ ，采样周期 =  $125 \mu\text{s}$ ，称为1帧(frame)
  - 在1帧时间里，可容纳的话路数有两种规格
    - 24路制:每帧由24个时间片(信道)和1位同步位组成
    - 30路制:每帧由32个时间片(信道)组成

# 波形编码(续)

## ◆ 2. 预测编码概念

- 预测编码(predictive coding)是对实际样本值与预测值之差进行编码的方法
- 在信道上传送的是相邻样本之间的差值，解码时使用差值和预测值进行重构，如图12所示

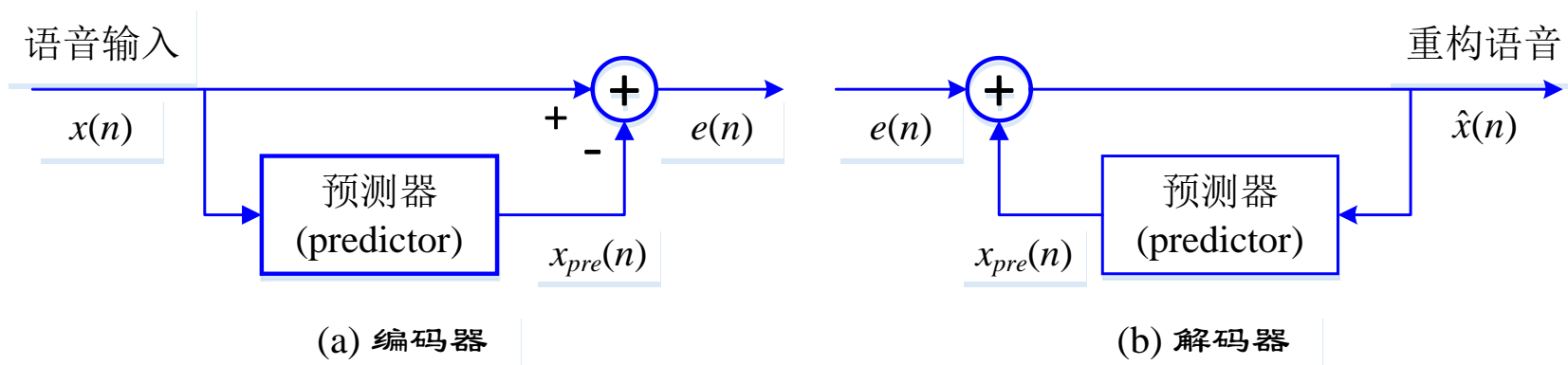


图12 预测编码框图

## 波形编码(续)

假设当前样本  $x(n)$  的预测值  $x_{pre}(n)$  为前一个样本值，预测误差  $e(n)$  可写成，

$$x_{pre}(n) = x(n-1)$$

$$e(n) = x(n) - x(n-1)$$

对差值  $e(n)$  编码比对原始语言样本编码，可用比较少的位数来表达一个样本。

对于具体的语音序列样本，如  $\cdots, x(n-2), x(n-1), x(n), \cdots$ ，如果使用过去几个样本值来预测当前的样本值，合理选择预测系数，产生的预测误差会更小。预测函数可写成，

$$x_{pre}(n) = \sum_{k=1}^N a_{n-k} x(n-k)$$

其中， $a_{n-k}$  为预测系数； $N$  为参加预测的样本数，通常  $N=2, 3$  或  $4$

**【例】** 假设  $x(n-3)=21$ ， $x(n-2)=24$ ， $x(n-1)=27$ ， $x(n)=23$ ，分别求  $N=1$  和  $N=2$  的预测误差  $e(n)$ 。

(1) 当  $N=1$  时， $e(n)=23-27=-4$ 。

(2) 当  $N=2$  时，为简单起见，假设预测系数  $a_{n-1} = a_{n-2} = 1/2$ ，那么

$$x_{pre}(n) = a_{n-1}x(n-1) + a_{n-2}x(n-2) = (27 + 24) / 2 \Rightarrow 25$$

$$e(n) = x(n) - x_{pre}(n) = 23 - 25 = -2$$



## 波形编码(续)

### ◆ 3. DPCM(差分脉冲编码调制)

- DCPM=differential pulse code modulation): 利用样本与样本之间存在冗余进行编码的数据压缩技术
- 基本思想: 根据过去的样本估算下一个样本信号幅度, 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

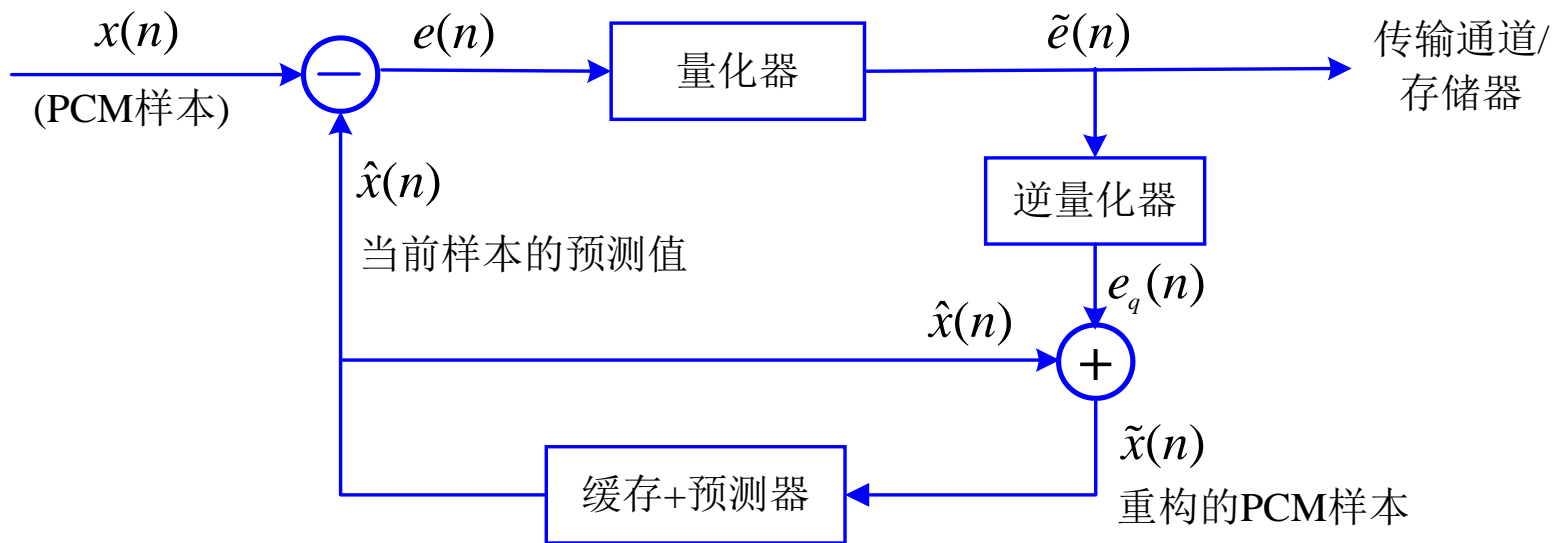
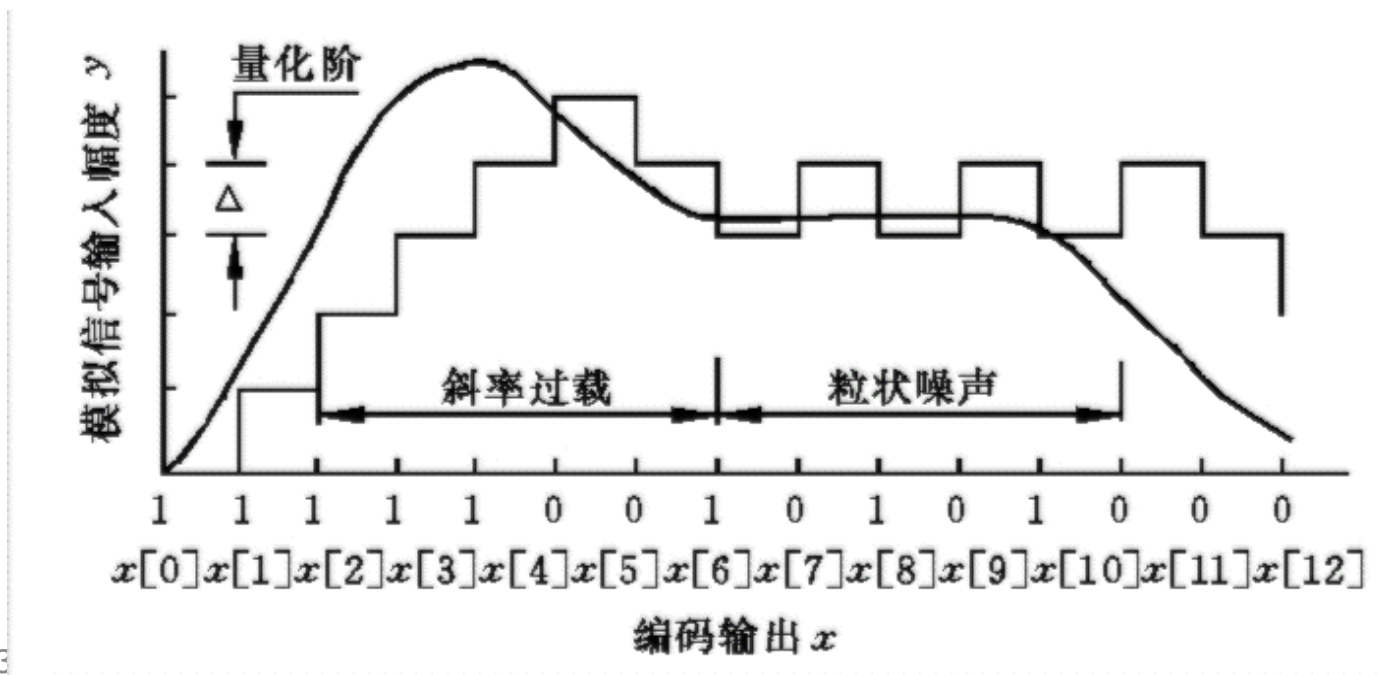


图13 DPCM方块图

## 波形编码(续)

### ◆ 4. DM (增量调制)

- 增量调制(delta modulation, DM): 是DPCM的简化版本, 常用作一个快速的模/数转换器
- 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, 差值为正用1表示, 差值为负用0表示
- 自适应增量调制(ADM): 根据输入信号斜率变化自动调整量化



## ◆ 5. ADPCM(自适应差分脉冲编码调制)

- ADPCM=adaptive difference pulse code modulation
- 综合APCM的自适应特性和DPCM的差分特性
- 核心理念
  - 利用自适应改变量化阶大小，用小的量化阶(step-size)编码小的差值，用大的量化阶编码大的差值
  - 使用过去的样本值估算当前输入样本的预测值，使实际样本值和预测值之间的差值总是最小
  - 编码和解码简化框图见图14

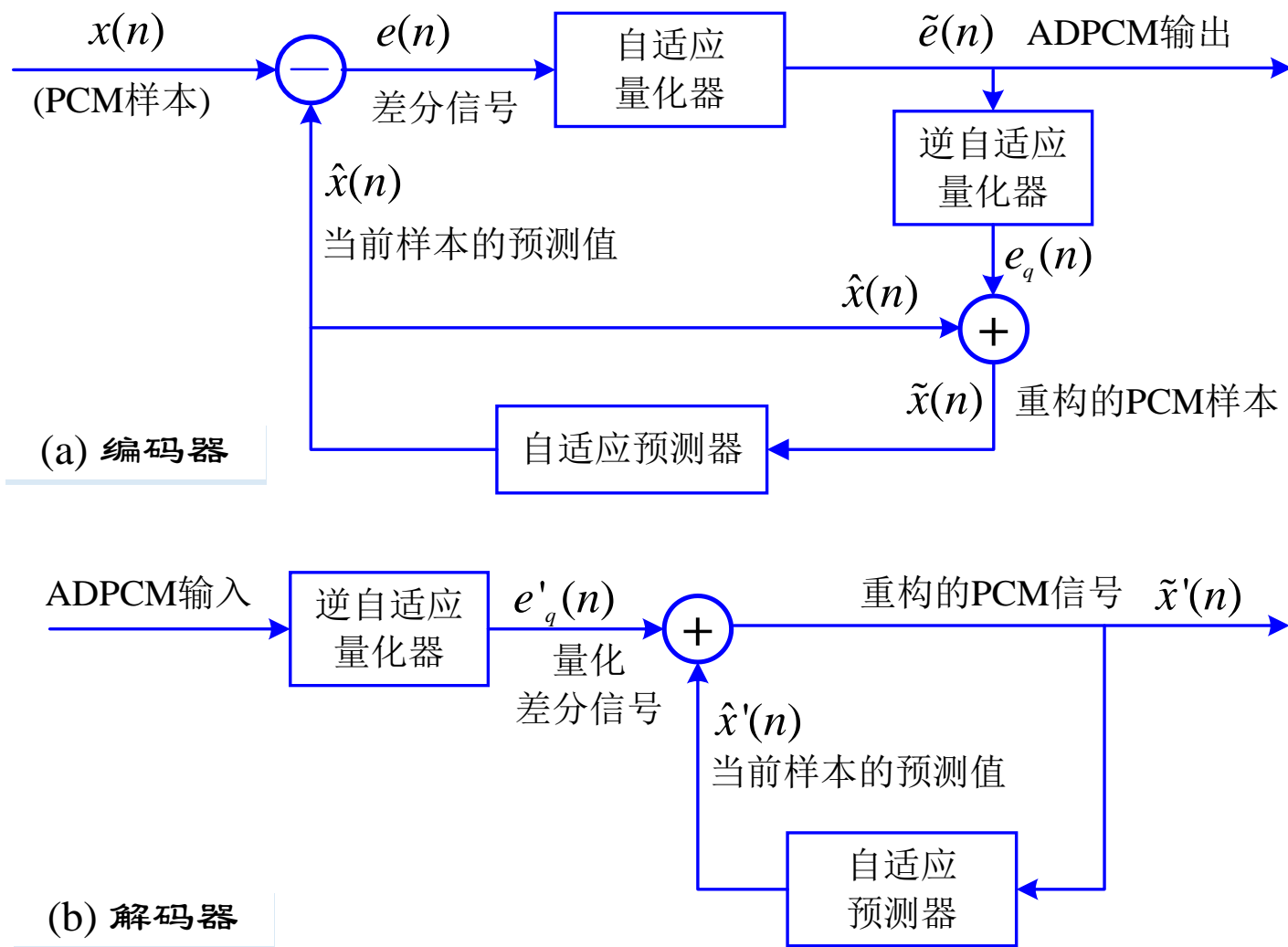


图14 ADPCM方块图

# MIDI:乐器数字接口

## 1. MIDI技术概述

### 什么是MIDI

MIDI是一种利用合成器产生合成音乐的技术。

Musical Instrument Digital Interface译为乐器数字接口。

利用数字信号处理技术合成的音效有：

- ①模仿钢琴、小提琴、吉他等音色；
- ②超越时空的太空音乐。

Midi示例：千千阙歌.mid



# MIDI:乐器数字接口(续)

## 2. MIDI标准

- ✓ MIDI是各种电子音乐设备之间以及与计算机交换信息的国际标准。(Yamaha, Roland)
- ✓ MIDI硬件规范：硬件接口标准和信号传输机制(I/O通道类型、连接电缆和插座形式)。
- ✓ MIDI软件规范：音乐信息数字化编码方式(音符、音符长短、音调和音量等)。

MIDI 音符部分代码表  
总计128个音符

编号	音符代码		所在的音阶	音调
	(二进制码)	(十六进制码)		
0	0000000	00	-1	C
1	0000001	01	-1	C#
2	0000010	02	-1	D
3	0000011	03	-1	D#
4	0000100	04	-1	E
5	0000101	05	-1	F

# MIDI:乐器数字接口(续)

## 3. MIDI合成方式

### 1. 调频合成法FM

原理：利用傅立叶级数进行分解。

特点：①系统开销小，声音较清脆；  
②声音音色少，音质差。

### 2. 波形表合成法WT

原理：预先存放真实乐器的声音样本

特点：①音乐真实感强，音色自然。  
②硬波形表（波形表存储在声卡的存储器中），软波形表。

# 数字音频的文件格式

## 1. WAV文件

- ✓ WAV是Microsoft/IBM共同开发的PC波形文件。因未经压缩，文件数据量很大。
- ✓ 通常使用WAV格式用来保存一些没有压缩的音频，也就是经过PCM编码后的音频，因此也称为波形文件，依照声音的波形进行存储，因此要占用较大的存储空间。
- ✓ CD唱片包含的就是WAVE格式的波形数据，只是扩展名为 “.cda”。
- ✓ 注意：WAV文件也可以存放压缩音频，但其本身的文件结构使之更加适合于存放原始音频数据并用作进一步的处理
- ✓ 特点：易于生成和编辑；在保证一定音质的前提下压缩比不大，不适合在网络上播放



# 数字音频的文件格式(续)



## 2. MP3文件

- ✓ MP3是MPEG 1 Layer III的简称，是由MPEG (Moving Pictures Expert Group)所制定的影音压缩—声音部分；
- ✓ MPEG声音部分，根据压缩质量和编码复杂程度划分为三层，即 Layer-1、Layer2、Layer3，且分别对应MP1、MP2、MP3这三种声音文件；
- ✓ MP3对音频信号采用的是有损压缩方式，高压缩比10:1-12:1；
- ✓ 为了降低声音失真度，MP3采取了「感官编码技术」——利用人耳听觉的特性，从声音中去除人耳听不到的信号；
- ✓ 保持良好的音质：利用人耳的特性，削减音乐中人耳听不到的成分，同时尝试尽可能的维持原来的声音，几乎达到了CD音质标准。

# 数字音频的文件格式(续)

## 3. WMA文件

- ✓ WMA (Windows Media Audio) 是Windows Media格式中的一个子集(音频格式)。
- ✓ 微软针对网络所开发的数字音乐文件格式
- ✓ 在只有64kbps的比特率情况下，可以达到接近CD的音质。
- ✓ 支援防复制功能，可以限制播放时间和播放次数甚至于播放的机器等等。
- ✓ 支持串流技术，即一边读一边播放，因此WMA可以很轻松的实现在线广播。
- ✓ 相对于MP3的主要优点是在较低的采样频率下音质要好些；缺点是算法相对复杂，且封闭。

# 数字音频的文件格式(续)

## 4. MIDI文件

- ✓ MIDI (乐器数字接口) 是由一组声音或乐器符号的集合;
- ✓ MIDI文件像记乐谱一样地记录下演奏的符号, 所以体积是所有音频格式中最小的;
- ✓ WAV文件记录声音数据, MIDI文件记录一系列乐谱指令;
- ✓ 编辑修改灵活方便, 可通过音序器自由的改变MIDI文件的曲调、音色、速度等, 甚至可以改换不同的乐器;
- ✓ MIDI声音仅适于重现打击乐或一些电子乐器的声音 (音源窄)。